ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Principles of Digital Communications: Summer Semester 2012 | Assignment date: Feb 29, 2012 | Due date: Feb 29, 2012

Solution of Homework 3

Problem 1. *(The "Wetterfrosch")*

(1) A and B must be chosen such that the suggested function become valid probability density functions, i.e. $\int_0^1 f_{Y|H}(y|i)dy = 1$ for $i = 0, 1$. This yields $A = 4/3$ and $B = 6/7$. (A quicker way is to draw the functions and find the area by looking at the drawings.)

(2) Let us first find the marginal of Y, i.e.

$$f_Y(y) = f_{Y|H}(y|0)p_H(0) + f_{Y|H}(y|1)p_H(1) = \frac{A+B}{2} + \frac{2B-3A}{12}y = \frac{23}{21} - \frac{4}{21}y,$$

Then, applying Bayes' rule gives

$$p_{H|Y}(0|y) = \frac{f_{Y|H}(y|0)p_H(0)}{f_Y(y)} = \frac{1}{2}\frac{A - \frac{A}{2}y}{\frac{23}{21} - \frac{4}{21}y},$$

and similarly

$$p_{H|Y}(1|y) = \frac{f_{Y|H}(y|1)p_H(1)}{f_Y(y)} = \frac{1}{2}\frac{B + \frac{B}{3}y}{\frac{23}{21} - \frac{4}{21}y}.$$

(3) The threshold is where the two functions are equal,

$$\frac{1}{2}\frac{A - \frac{A}{2}y}{\frac{23}{21} - \frac{4}{21}y} = \frac{1}{2}\frac{B + \frac{B}{3}y}{\frac{23}{21} - \frac{4}{21}y},$$

or equivalently,

$$A - \frac{A}{2}y = B + \frac{B}{3}y.$$

The y that satisfies this equation is our threshold θ, thus

$$\theta = \frac{A - B}{\frac{B}{3} + \frac{A}{2}} = 0.5.$$

(4) The probability that we decide $\hat{H}(y) = 1$ when in reality, $H = 0$, is just the probability that y is *larger* than the threshold, given that $H = 0$, which is

$$\begin{aligned} Pr(Y > \theta | H = 0) &= \int_\theta^1 f_{Y|H}(y|0) dy = \int_\theta^1 \left(A - \frac{A}{2} y \right) dy \\ &= (1-\theta)A - \frac{A}{2}\frac{1-\theta^2}{2}. \end{aligned}$$

(5) By analogy to *(4)*,

$$\begin{aligned} Pr(Y < \theta | H = 1) &= \int_0^\theta f_{Y|H}(y|1) dy = \int_0^\theta \left(B + \frac{B}{3} y \right) dy \\ &= \theta B + \frac{B}{3}\frac{\theta^2}{2}. \end{aligned}$$

So we find

$$Pr(\text{wrong}) = \frac{1}{2}\left((1-\theta)A - \frac{A}{2}\frac{1-\theta^2}{2} + \theta B + \frac{B}{3}\frac{\theta^2}{2} \right).$$

(6) To minimize $Pr(\text{wrong})$ over θ, we take the derivative with respect to θ, i.e.

$$\frac{d}{d\theta} Pr(\text{wrong}) = \frac{1}{2}\left(-A - \frac{A}{2}\frac{-2\theta}{2} + B + \frac{B}{3}\frac{2\theta}{2} \right) = \frac{1}{2}\left(-A + \frac{A}{2}\theta + B + \frac{B}{3}\theta \right).$$

Setting this equal to zero, we find

$$\theta = \frac{A - B}{\frac{B}{3} + \frac{A}{2}} = 0.5,$$

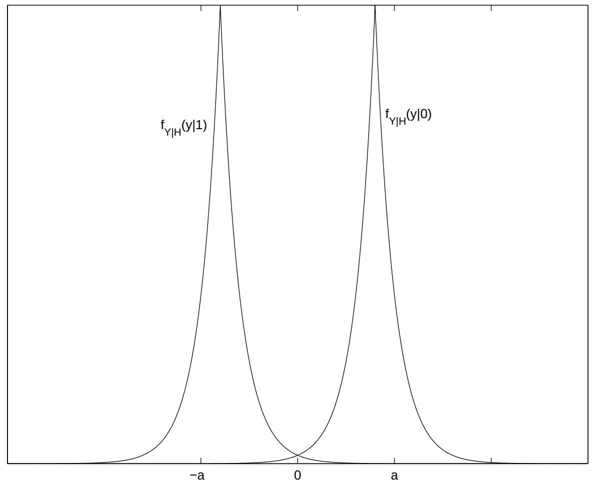
which verifies that the MAP decision rule (as derived in *(3)*) minimizes the probability of erroneous weather forecast.

Problem 2. *(Hypothesis Testing in Laplacian Noise)*

(1) We find the following conditional densities for the observation Y under hypothesis $H = 0$ and $H = 1$, respectively:

$$\begin{aligned} f_{Y|H}(y|0) &= \frac{1}{2} e^{-|y-a|} \\ f_{Y|H}(y|1) &= \frac{1}{2} e^{-|y+a|}. \end{aligned}$$

(2) Note that $f_{Y|H}(y|0)$ is $f_Z(y-a)$ and $f_{Y|H}(y|1)$ is $f_Z(y+a)$. From a picture of $f_{Y|H}(y|0)$ and $f_{Y|H}(y|0)$ we see immediately that a maximum likelihood decision rule decides for $H = 0$ when $y > 0$ and for $H = 1$ when $y < 0$.



(3)

$$\begin{aligned} Pr\{e|H=0\} &= Pr(y<0|H=0) = \int_{-\infty}^{0} f_{Y|H}(y|0)dy \\ &= \int_{-\infty}^{0} \frac{1}{2} e^{-|y-a|} dy = \int_{-\infty}^{0} \frac{1}{2} e^{(y-a)} dy \\ &= \frac{e^{-a}}{2}\, e^{y}|_{-\infty}^{0} = \frac{e^{-a}}{2}. \end{aligned}$$

By symmetry, we find that

$$Pr\{e|H=1\} = \frac{e^{-a}}{2},$$

and thus,

$$Pr\{e\} = Pr\{e|H=0\}p_H(0) + Pr\{e|H=1\}p_H(1) = \frac{e^{-a}}{2}.$$

Problem 3. *(Discrete additive Gaussian channel)*

1. We define two hypotheses :

$$\begin{cases} H_0 : \text{we send } 0 \\ H_1 : \text{we send } 1. \end{cases}$$

Under

$$\begin{cases} H_0 : R_i \sim N(-A, \sigma^2) \text{ and } P(Y_1, ..., Y_n|H_0) = \dfrac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\sum_{i=1}^{n} \frac{(Y_i+A)^2}{2\sigma^2}} \\ H_1 : R_i \sim N(A, \sigma^2) \text{ and } P(Y_1, ..., Y_n|H_1) = \dfrac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\sum_{i=1}^{n} \frac{(Y_i-A)^2}{2\sigma^2}} \end{cases}$$

2. We find the likelihood ratio

$$L(Y_1, ..., Y_n) = \frac{P(Y_1, ..., Y_n|H_1)}{P(Y_1, ..., Y_n|H_0)} = e^{\frac{2A}{\sigma^2}\sum_{i\leq 1}^{n} y_i}$$

   And the MAP rule is

$$L(Y_1, ..., Y_N) \underset{H_1}{\overset{H_0}{\lessgtr}} 1$$

   As $log(.)$ is an increasing function we can take the logarithm on both sides. Hence we obtain

$$\sum_{i=1}^{n} y_i \underset{H_0}{\overset{H_1}{\lessgtr}} 0$$

3. Under H_1 , $\sum_{i=1}^{n} y_i \sim N(nA, n\sigma^2)$. Hence

$$\begin{aligned} P(E|H_1) &= P\left(\sum_{i=1}^{n} y_i < 0|H_1\right) \\ &= Q\left(\frac{nA}{\sqrt{n\sigma^2}}\right) \\ &= Q\left(\sqrt{n\frac{A^2}{\sigma^2}}\right) \\ &= Q(\sqrt{nSNR}) \end{aligned}$$

   Similarly we can show that $P(E|H_0) = Q(\sqrt{nSNR})$. Hence $P(E) = Q(\sqrt{nSNR})$.

4. $P(E) = Q(\sqrt{nSNR}) \leq \frac{1}{2}e^{-n\frac{SNR}{2}}$, which goes exponentially fast to zero as a function of n.

Problem 4. *(Poisson Parameter Estimation)*

(1) We can write the MAP decision rule in the following shape:

$$\frac{p_{Y|H}(y|1)}{p_{Y|H}(y|0)} \underset{0}{\overset{1}{\gtrless}} \frac{p_H(0)}{p_H(1)}. \tag{1}$$

Plugging in, we find

$$\frac{\lambda_1^y e^{-\lambda_1}}{\lambda_0^y e^{-\lambda_0}} \underset{0}{\overset{1}{\gtrless}} \frac{p_0}{1-p_0}, \tag{2}$$

and then

$$\left(\frac{\lambda_1}{\lambda_0}\right)^y \underset{0}{\overset{1}{\gtrless}} \frac{p_0}{1-p_0} e^{\lambda_1-\lambda_0}. \tag{3}$$

Taking logarithms on both sides does not change the direction of the inequalities, therefore

$$y \log\left(\frac{\lambda_1}{\lambda_0}\right) \underset{0}{\overset{1}{\gtrless}} \log\left(\frac{p_0}{1-p_0} e^{\lambda_1-\lambda_0}\right). \tag{4}$$

Now comes the critical step: the term $\log(\lambda_1/\lambda_0)$ can be negative, and if it is, then dividing by it involves changing the direction of the inequality. This issue is discussed in more detail in the next subsection.

(2) Suppose for the moment $\lambda_1 > \lambda_0$. Then, $\log(\lambda_1/\lambda_0) > 0$, and the decision rule becomes

$$y \quad \overset{1}{\underset{0}{\gtrless}} \quad \frac{\log\left(\frac{p_0}{1-p_0}e^{\lambda_1-\lambda_0}\right)}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \quad \stackrel{\text{def}}{=} \quad \theta. \tag{5}$$

We compute

$$\begin{aligned} Pr\{e|H=0\} \quad &= \quad Pr(y > \theta|H=0) = \sum_{y=\lceil\theta\rceil}^{\infty} p_{Y|H}(y|0) &(6)\\ &= \quad 1 - \sum_{y=0}^{\lfloor\theta\rfloor} \frac{\lambda_0^y}{y!}e^{-\lambda_0}, &(7)\end{aligned}$$

and by analogy

$$\begin{aligned} Pr\{e|H=1\} \quad &= \quad Pr(y < \theta|H=1) = \sum_{y=0}^{\lfloor\theta\rfloor} p_{Y|H}(y|1) &(8)\\ &= \quad \sum_{y=0}^{\lfloor\theta\rfloor} \frac{\lambda_1^y}{y!}e^{-\lambda_1}. &(9)\end{aligned}$$

Thus, the probability of error becomes

$$Pr\{e\} \quad = \quad p_0\left(1 - \sum_{y=0}^{\lfloor\theta\rfloor} \frac{\lambda_0^y}{y!}e^{-\lambda_0}\right) + (1-p_0)\sum_{y=0}^{\lfloor\theta\rfloor} \frac{\lambda_1^y}{y!}e^{-\lambda_1}. \tag{10}$$

Now, suppose that $\lambda_1 < \lambda_0$. Then, $\log(\lambda_1/\lambda_0) < 0$, and we have to swap the inequality sign, thus

$$y \quad \overset{1}{\underset{0}{\lessgtr}} \quad \frac{\log\left(\frac{p_0}{1-p_0}e^{\lambda_1-\lambda_0}\right)}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \quad \stackrel{\text{def}}{=} \quad \theta. \tag{11}$$

The rest of the analysis goes along the same lines, and finally, we obtain

$$Pr\{e\} \quad = \quad p_0\sum_{y=0}^{\lfloor\theta\rfloor} \frac{\lambda_0^y}{y!}e^{-\lambda_0} + (1-p_0)\left(1 - \sum_{y=0}^{\lfloor\theta\rfloor} \frac{\lambda_1^y}{y!}e^{-\lambda_1}\right). \tag{12}$$

The case $\lambda_0 = \lambda_1$ yields $\log(\lambda_0/\lambda_1) = 0$, so the decision rule becomes $0 \overset{1}{\underset{0}{\gtrless}} \theta$, independent of the observation y. Thus, we may exclude the case $\lambda_0 = \lambda_1$ from our discussion.

(3) Here, we are in the case $\lambda_1 > \lambda_0$, and we find $\theta \approx 4.54$. We thus evaluate

$$Pr\{e\} = \frac{1}{3}\left(1 - \sum_{y=0}^{4} \frac{2^y}{y!} e^{-2}\right) + \frac{2}{3} \sum_{y=0}^{4} \left(\frac{10^y}{y!} e^{-10}\right) \approx 0.1472. \tag{13}$$

(4) We find $\theta \approx 7.5163$

$$Pr\{e\} = \frac{1}{3}\left(1 - \sum_{y=0}^{7} \frac{2^y}{y!} e^{-2}\right) + \frac{2}{3} \sum_{y=0}^{7} \left(\frac{20^y}{y!} e^{-20}\right) \approx 0.000885. \tag{14}$$

The two Poisson distributions are much better separated than in *(3)*; therefore, it becomes considerably easier to distinguish them based on one single observation y.

Problem 5. *(IID versus First-Order Markov)*

An explanation regarding the title of this problem: *i.i.d.* stands for: independent and identically distributed; this means that all the observations $Y_1, \ldots, Y_k$ have the same probability mass function and are independent of each other. *First-order Markov* means that the observations $Y_1, \ldots, Y_k$ depend on each other in a particular way: the probability mass function of observation Y_i depends on the value of Y_{i-1}, but given the value of Y_{i-1}, it is independent of the earlier observations $Y_1, \ldots, Y_{i-2}$. Thus, in this problem, we observe a binary sequence, and we want to know whether it has been generated by an *i.i.d.* source or by a first-order Markov source.

1. Since the two hypotheses are equally likely, we find

   $$L(y) \stackrel{\text{def}}{=} \frac{f_{Y|H}(y|1)}{f_{Y|H}(y|0)} \underset{0}{\overset{1}{\gtrless}} \frac{p_H(0)}{p_H(1)} = 1.$$

   Plugging in,

   $$\frac{1/2 \cdot (1/4)^l \cdot (3/4)^{k-l-1}}{(1/2)^k} \underset{0}{\overset{1}{\gtrless}} 1.$$

   where l is the number of times the observed sequence changes either from zero to one or from one to zero, i.e. the number of transitions in the observed sequence.

2. The sufficient statistic here is simply the number of transitions l; this entirely specifies the likelihood ratio. This means that when observing the sequence y, all you have to remember is the number of transitions; the rest can be thrown away immediately.

   We can write the sufficient statistic as

   $$T(y) = \text{Number of transitions in the observed sequence.}$$

   The irrelevant data $h(y)$ would be the actual location of these transitions: These locations do not influence the MAP hypothesis testing problem. For the functions

$g_0(\cdot)$ and $g_1(\cdot)$, (refer to the problem on sufficient statistics at the end of chapter 2) we find

$$\begin{aligned} g_0(T(y)) &= \left(\frac{1}{2}\right)^k \\ g_1(T(y)) &= \frac{1}{2}\frac{1}{4}^{T(y)}\frac{3}{4}^{k-T(y)-1}. \end{aligned}$$

3. So, in this case, the number of non-transitions is $(k-l)=s$, and the log-likelihood ratio becomes

$$\begin{aligned} \log\frac{1/2\cdot(1/4)^{k-s}\cdot(3/4)^{s-1}}{(1/2)^k} &= \log\frac{(1/4)^{k-s}\cdot(3/4)^{s-1}}{(1/2)^{k-1}} \\ &= (k-s)\log(1/4)+(s-1)\log(3/4)-(k-1)\log(1/2) \\ &= s\log\frac{3/4}{1/4}+k\log\frac{1/4}{1/2}+\log\frac{1/2}{3/4}. \end{aligned}$$

Thus, in terms of this log-likelihood ratio, the decision rule becomes

$$s\log\frac{3/4}{1/4}+k\log\frac{1/4}{1/2}+\log\frac{1/2}{3/4} \underset{0}{\overset{1}{\gtrless}} 0.$$

That is, we have to find the smallest possible s such that this expression becomes larger or equal to zero. This is

$$s = \left\lceil \frac{k\log\frac{1/4}{1/2}+\log\frac{1/2}{3/4}}{\log\frac{1/4}{3/4}} \right\rceil.$$

Plugging in $k=20$, we obtain that the smallest s is $s=13$.

Note: Do you understand why it does *not* matter which logarithm we use as long as we use the same logarithm everywhere?

Problem 6. *(One Bit over a Binary Channel with Memory)*

(1) From the receiver operation we see that Y_1 is available and is equal to $\hat{Y}_1$. Now $\hat{Y}_2 = Y_2 \oplus Y_1$ but now we know Y_1 and if we calculate module 2 sum of Y_1 and $\hat{Y}_2$ we will have $\hat{Y}_2 + Y_1 = Y_2 \oplus Y_1 \oplus Y_1 = Y_2$ and so we can recover Y_2 from $\hat{Y}_1$ and $\hat{Y}_2$ and continuing this way at stage m we have computed $Y_1, Y_2, \cdots, Y_m$ from $\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_m$ and now if we compute $\hat{Y}_{m+1} \oplus Y_m$ we will obtain Y_{m+1} and so we can recover all of the sequence of Y from sequence $\hat{Y}$ and so sequence Y is a function of sequence $\hat{Y}$ but sequence $\hat{Y}$ is a function of sequence Y by construction and so these two sequences are equivalent and so $(\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_n)$ is a sufficient statistic for $(Y_1, Y_2, \cdots, Y_n)$.

(2) For the bit 0 the sequence we sent is $(X_1^{(0)}, \cdots, X_n^{(0)})$ so for the sequence $\hat{Y}$ we have $\hat{Y}_1 = Y_1 = X_1^{(0)} + Z_1$ and $\hat{Y}_i = Y_i \oplus Y_{i-1} = X_i^{(0)} \oplus X_{i-1}^{(0)} \oplus Z_i \oplus Z_{i-1}$ but from noise construction method we see that $Z_i \oplus Z_{i-1} = N_i$ so simplifying the result we will have $\hat{Y}_i = X_i^{(0)} + X_{i-1}^{(0)} + N_i$. Using a similar method for bit 1 we have $\hat{Y}_1 = X_1^{(1)} + Z_1$ and $\hat{Y}_i = X_i^{(1)} \oplus X_{i-1}^{(1)} + N_i$ which holds for $i = 2, 3, \cdots, n$.

(3) First of all we see that the effective processing of channel and receiver results in a channel $X \to \hat{Y}$ in which transmitted sequence $(X_1, X_2, \cdots, X_n)$ is transformed to sequence $(X_1, X_2 \oplus X_1, \cdots, X_n \oplus X_{n-1})$ at the output of the receiver processor. As can be seen from the part *(2)* the noise term added to the output sequence is $(Z_1, N_2, \cdots, N_n)$. We know that Z_1 is independent of N_i and has the same distribution as N_i so noise term is a sequence of *iid* binary random variable so it behaves similar to n independent uses of binary symmetric channel. For the binary symmetric channel we know that in order to minimize the probability of the error transmitted sequences under two hypotheses should differ from each other in every term for example all zero and all one sequences $(0, 0, \cdots, 0)$ and $(1, 1, \cdots, 1)$ have this property. In this case the signal term in the received sequence is $(X_1, X_2 \oplus X_1, \cdots, X_n \oplus X_{n-1})$ so we should choose X_i such that signal terms has the maximum distance under different hypotheses. It can be showed that alternating 0 and 1 sequences $(1, 0, 1, 0, \cdots)$ for H_1 and all 0 sequence $(0, 0, 0, 0, \cdots)$ for H_0 result in maximal distance because in the former, at the output of the receiver processor, we will reach to all one sequence $(1, 1, 1, 1, \cdots)$ and in the latter to all zero sequence $(0, 0, \cdots, 0)$ which have the maximum mutual distance and result in minimum probability of error in MAP post-processor.