

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

LECTURE NOTES IN

Principle of Digital Communications



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

OLIVIER CLOUX

SPRING 2017



Contents

1 Introduction	1
1.1 Probability review	1
2 Lecture 2	3
2.1 Hypothesis testing	3
3 Lecture 3	4

1 Introduction

Communications are basically going through a medium with some random noise, that distort the signal. From an emitted signal $x(t)$ and noise $N(t)$, we obtain the result $r(t)$ by some composition of both. To avoid ending up with a fucked up signal, we create “boxes” : before sending an encoder and an waveform (bits -> waves) ; after sending, on the other hand, we design a decoder.

The encoder should take k bits, and output $n \geq k$ bits to add some redundancy. Hence, even if some bits are corrupted, we can reconstruct the signal and the decoder can find the original signal perfectly. The waveform box will simple take these n bits to output 2D signal.

1.1 Probability review

Let $\omega \in \Omega$ be the outcome of a “random experiment”.

Example 1. Let $\Omega = [0, 1]$, and let ω be uniformly distributed in Ω . Then for

$$X(\omega) = \begin{cases} 0 & 0 \leq \omega < \frac{1}{2} \\ 1 & \frac{1}{2} < \omega \leq 1 \end{cases}$$

Then $P(X = 0) = P(X = 1) = \frac{1}{2}$

For the same support, we define $Y(\omega)$ as 0 when ω is in $\{0, \frac{1}{4}\} \cup \{\frac{1}{2}, \frac{3}{4}\}$, 1 elsewhere.
And $Z(\omega)$ as 1 between $\frac{1}{4}$ and $\frac{3}{4}$, 0 elsewhere. The of course,

$$P(Y = 1) = P(Y = 0) = P(X = Y) = \frac{1}{2}$$

while for every $x, y, z \in \{0, 1\}$

$$P((X, Z) = (x, z)) = P((X, Y) = (x, y)) = P((Y, Z) = (y, z)) = \frac{1}{4}$$

We recall the definition of *independence* :

Definition 1. X, Y are **independent** if

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

And for 3 variables :

Definition 2. X, Y, Z are independent if

$$P((X, Y, Z) = (x, y, z)) = P(X = x)P(Y = y)P(Z = z)$$

Let's review also the definition of *conditional probabilities*. The probability for a variable, having knowledge of the other. It is written as

$$P_{Y|X}(y|x) \equiv P(Y = y|X = x)$$

From that, we know have more definitions for independence :

Definition 3. X, Y are independent if and only if

$$P(Y = y|X = x) = P(Y = y)$$

If they are independent, having knowledge on one has no effect on the second.

The following rule (Bayes') is also very useful :

Definition 4. Bayes' rule :

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

Following, the Condition Independence. X, Y are independent, condition on Z (written $X - Z - Y$) means, for all x, y, z :

$$\begin{aligned} P(X = x, Y = y | Z = z) &= P(X = x | Z = z) P(Y = y | Z = z) \\ &= P(X = x, Y = y) P(Z = z) \\ &= P(X = x, Z = z) P(Y = y, Z = z) \end{aligned}$$

Expected value :

$$E[f(X)] = \sum_x f(x) P(X = x)$$

If (X, Y) are independent, then $E[XY] = E[X]E[Y]$

The Law of large numbers : Suppose X_1, X_2, X_3, \dots are independent, \mathbb{R} -valued random variables such as $P(X_n = x) = p(x)$. Then $A_N = \frac{1}{N} \sum_{i=1}^N \xrightarrow[N \rightarrow \infty]{} E[X]$

2 Lecture 2

A sender has a message in mind, and wants to send it. The message (i) is not completely arbitrary, and belongs in a set \mathcal{H} (*the set of possible messages*). We pass the message through a *transmitter* that changes the message to a signal c_i . This c_i belongs in some vector space ; for now, we say this vector space is \mathbb{R}^n . This vector is then sent through a channel to obtain some $y \in \mathbb{R}^n$. But the channel has such properties that y can't be predicted perfectly. We need then to analyse the probability that, given the original signal c as input to the channel, we obtain this particular y . Then, y is going through a *receiver* that will give an estimate of i .

2.1 Hypothesis testing

Let's start with a simple set $\mathcal{H} = \{0, 1\}$. We know that $P(H = 0) = 0.7$, while $P(H = 1) = 0.3$. Our guess for H is written \hat{H} .

Then, our goal is to maximize $P(\hat{H} = H)$. The most logical solution, if there is no observation at all, is to guess $\hat{H} = \operatorname{argmax}_{i \in \mathcal{H}} P(H = i)$. But now suppose we have

$$P_{Y|H}(y|i)$$

That is the probability that the observation is y when the hypothesis is i . Once gone through the channel, we have $y \in \{a, b\}$, with repartition ($P(Y = a|H = 0) = 0.9$)

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.6 & 0.5 \end{bmatrix}$$

Suppose we observe $y = a$. We look for

$$\begin{aligned} P(H = 0|Y = a) &= \frac{P(H = 0, Y = a)}{P(Y = a)} = \frac{P(H = 0)P(Y = a|H = 0)}{P(Y = a)} = \frac{0.7 \cdot 0.9}{P(Y = a)} \\ P(H = 1|Y = a) &= \frac{P(H = 1)P(Y = a|H = 1)}{P(Y = a)} = \frac{0.3 \cdot 0.6}{P(Y = a)} \end{aligned}$$

Then logically, $\hat{H} = 0$ as the first probability is much bigger.

Now suppose we observe that $Y = b$; after applying the same computations as above, we find $P(H = 0|Y = b) = \frac{0.7 \cdot 0.1}{\dots}$ and $P(H = 1|Y = b) = \frac{0.3 \cdot 0.4}{\dots}$, and thus $\hat{b} = 1 = 1$.

Definition 5. This approach is known as the **MAP rule** (max aposteriori proh)

3 Lecture 3

Reminder about the binary hypothesis testing. We have an hypothesis $H \in \{0, 1\}$, that influences an observation y . Based on this observation, we make a decision $\hat{H} \in \{0, 1\}$. We can't decide "not to decide", we must chose something and thus will be either wrong or right. Thus, the system ($H - Y - \hat{H}$) is Markov, and is

$$P(H = h, \hat{H} = \hat{h}|Y = y) = P(H = h|Y = y)P(\hat{H} = \hat{h}|Y = y)$$

As the decision is completely decided on y , we define

$$P(\hat{H} = 1|Y = y) = \phi(y)$$

Note that sometimes, if we are undecided, we might "toss a coin" to decide. This model allows randomized decision strategies.

Likelihood ratio test : those are rules of the type that for some “threshold” t , $\phi(y)$ is of the form

$$\phi(y) = \begin{cases} 1 & \Lambda(y) = \frac{P(Y=y|H=1)}{P(Y=y|H=0)} > t \\ 0 & < t \end{cases}$$

Given a rule, we can compute $P(\hat{H} \neq H)$, that is the probability of error. Sometimes, it is more interesting to compute a second number $P(\hat{H} = 1|H = 0)$ (and the other way around $P(\hat{H} = 0|H = 1)$) ; those are the two error we can make. They are defined

$$P(\hat{H} = 1|H = 0) = \sum_y P(Y = y|H = 0)\phi(y) \quad (1)$$

$$P(\hat{H} = 0|H = 1) = \sum_y P(Y = y|H = 1)(1 - \phi(y)) \quad (2)$$

$$(3)$$

Consider $tP(\hat{H} = 1|H = 0) + P(\hat{H} = 0|H = 1) = \sum_y \phi(y) \underbrace{tP(Y=y|H=0)}_{P(Y=y|H=0)} + (1 - \phi(y)) \underbrace{P(Y=y|H=1)}_{P(Y=y|H=1)}$ And this is $\geq \sum_y \min\{tP(Y=y|H=0), P(Y=y|H=1)\}$

Lemma : The likelihood ratio test with threshold t minimizes (among all possible tests ϕ) $tP(\hat{H} = 1|H = 0) + P(\hat{H} = 0|H = 1)$

For the likelihood ratio test, we have

$$\phi(y) = \begin{cases} 1 & tP(Y = y|H = 0) < P(Y = y|H = 1) \\ 0 & tP(Y = y|H = 0) > P(Y = y|H = 1) \end{cases}$$

Lemma (Neymann-Pierson) : For any $0 \leq \alpha \leq 1$, there is a likelihood ratio test ϕ with $\alpha = P_\phi(\hat{H} = 1|H = 0)$.

For a second claim : For any other test ψ , either $P_\psi(\hat{H} = 1|H = 0) \geq P_\phi(\hat{H} = 1|H = 0)$ or $P_\psi(\hat{H} = 0|H = 1) \geq P_\phi(\hat{H} = 0|H = 1)$