

Computer Networking and IT Security (INHN0012)

Tutorial 2

Problem 1 Source entropy

As we will see in the lecture, a message source can be abstractly modeled as follows:

- A message source Q emits statistically independent character from an alphabet $\mathcal{X} = \{a, b\}$.
- The source Q is assumed to be memoryless, i. e., the output in step $n >$ does not depend on the output of any previous step.
- The emitted symbols a, b carry a different amount of information, depending on their probability of occurrence – in general, the less likely a symbol is being emitted, the higher the information content.
- Information measured in bit

We model this message source as a discrete random variable X . The probability that the source emits the character $X = a$ is $p_a = \Pr[X = a] = 0.25$.

a)* Determine the probability p_b that the character $X = b$ is emitted.

Since $p_a + p_b = 1$, it follows that $p_b = 0.75$.
--

b) Determine the information content $I(a)$ and $I(b)$ of both symbols.

$I(a) = -\log_2 p_a = 2.00 \text{ bit}$ $I(b) = -\log_2 p_b \approx 0.42 \text{ bit}$
--

c) Determine the entropy H of the source.

$H(X) = \sum_{x \in \mathcal{X}} p_x I(x) = 0.81 \text{ bit/symbol}$
--

- d) Determine the occurrence probabilities p_0 and p_1 of another binary message source Q' such that its entropy H is maximal.

First we express p_1 in dependence of p_0 and write $p_1 = 1 - p_0$. For simplicity we write $p_0 = p$. Finally, the entropy H can be expressed as a function of p and the probability we are looking for can be determined by means of the derivative.

$$H = -p \log_2(p) - (1 - p) \log_2(1 - p)$$
$$\frac{dH}{dp} = -\log_2(p) - \frac{p}{p \ln(2)} + \log_2(1 - p) + \frac{1 - p}{(1 - p) \ln(2)}$$
$$\Rightarrow \log_2(p) + \frac{p}{p \ln(2)} \stackrel{!}{=} \log_2(1 - p) + \frac{1 - p}{(1 - p) \ln(2)}$$

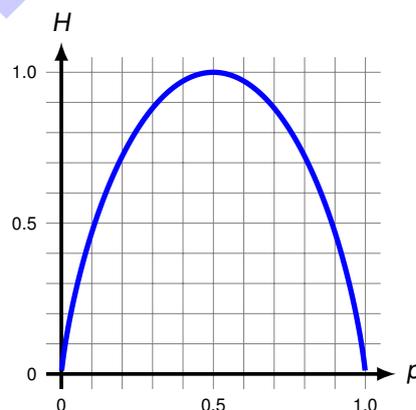
Comparing both sides yields $p = 1 - p = 1/2$.

- e) What is the maximum entropy of a binary source?

The entropy is maximised when $\Pr[X = a] = \Pr[X = b] = 0.5$ holds. The maximum entropy is therefore

$$H_{\max} = -2 \cdot 0.5 \cdot \log_2(0.5) = 1 \text{ bit/symbol.}$$

- f) Sketch the source entropy H of a binary source in general as a function of the probability of occurrence p .



g) Obviously, the entropy $H(X) < 1$ is not maximal. What conclusion can be drawn from this fact for the data stream emitted by source Q with respect to redundancy?

The string emitted by Q , which is nothing but different instantiations of the random variable X , contains redundancy. The data stream generated by Q is therefore representable on average with less than 1 bit/symbol.

h) Generalise the results of the subtasks d) and e) to a source emitting n different symbols.

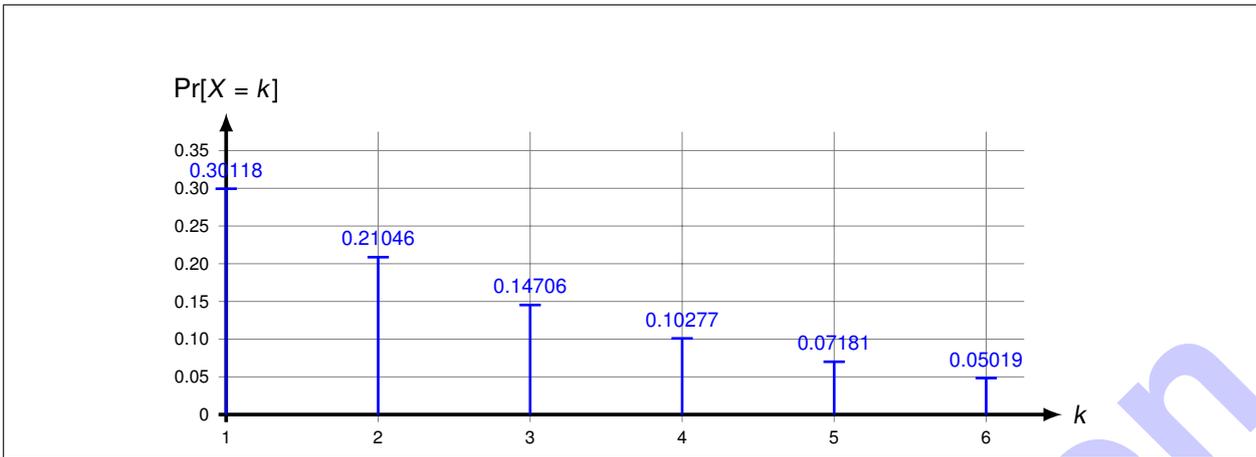
In general, the following applies to entropy

$$H(X) = \sum_{x \in \mathcal{X}} I(x)p_i.$$

With the requirement $p_i = p$, i.e. all characters occur with the same probability, it immediately follows $p = 1/N$ and thus

$$H = \sum_{x \in \mathcal{X}} I(x)p = - \sum_{i=1}^N \log_2 \left(\frac{1}{N} \right) \frac{1}{N} = \log_2(N).$$

d)* Sketch the probability from subtask c) for $k \in \{1, \dots, 6\}$.



e) Assume that the responsible protocol on the link layer aborts the retransmission if the third transmission attempt was unsuccessful. What is the probability that a frame cannot be transmitted?

The probability corresponds to the transmission failing three times in a row without regard to whether it works the 4th time or not. This results to

$$\Pr[X > 3] = 1 - \Pr[X \leq 3] = 1 - \sum_{k=1}^3 \Pr[X = k] \approx 34\%$$

Alternative solution:

$$\Pr[X > 3] = (1 - p_R)^3 \approx 34\%$$

Attention: The alternative solution is only correct because X is geometrically distributed and the geometric distribution is memoryless, i.e. the failure of the k -th transmission does not influence the $(k + 1)$ -th transmission. If this independence were not fulfilled, the alternative solution would give a wrong result!

Sample Solution