Quantum Information Theory

PD Dr. Joseph M. Renes

Winter Semester 2012/2013

TECHNISCHE UNIVERSITÄT DARMSTADT

Problem Set #5 **13 March 2013**

Problem 5.1 Information and Description Length

Suppose we have a collection of 12 balls, all identical except that one is either lighter or heavier than the rest. At our disposal is a two-pan balance onto which we can place any number of balls in the left pan and the same number in the right pan. The balance registers which of the pans is heavier, or that they are equal. What is the fewest number of weighings needed to determine the odd ball and whether it is lighter or heavier? It might be useful to consider the following questions:

a) How much information is gained upon learning (i) the state of a flipped coin; (ii) the states of two flipped coins; (iii) the outcome of the roll of a four-sided die?

Solution: Assuming the coin is fair, we learn one bit from one flipped coin, two bits from two coins, and also two bits from the (fair) four-sided die. If the states are equally-likely, the entropy is the log of the number of them.

b) How much information is gained when the odd ball and its weight are identified?

Solution: There are 24 possible states as any of the 12 balls could be either light or heavy. Hence we learn $\log 24$ bits.

c) How much information is gained on the first step if six balls are weighed against the other six? How much is gained by first weighing four against another four, leaving the rest aside?

Solution: Weighing six against six rules out half of the possibilities. Suppose we weigh 1-6 versus 7-12. If the 1-6 pan comes up lighter, then there are only 12 possibilities remaining: either one of 1-6 is light or one of 7-12 is heavy. The opposite holds for the opposite result, so we've shrunk the list of possible states from 24 to 12, an information gain of one bit.
In contrast, weighing two arbitrary sets of four balls gives more information. Suppose we weigh 1-4 versus 5-8. If they balance, then one of 9-12 is the odd ball and is either light or heavy, for a total of 8 possibilities. If 1-4 is lighter, then either one of 1-4 is light or one of 5-8 is heavy, again 8 possibilities. The opposite states are possible if 1-4 is heavy. Therefore, no matter the outcome the list of possibilities has shrunk from 24 elements to 8, a gain of $\log 3$ bits.

d) For your prospective weighing strategy, draw a tree showing the possible outcomes of the chosen weighing and what weighing is to be performed next. At each node, how much information has been gained and how much remains to be gained?

Solution: The goal in designing a strategy is to maximize the information gain at every step. Since there are 24 possible states and each weighing could in principle have three equally-likely outcomes, it will take at least 3 weighings to determine which ball is odd and its weight ($3^2 < 24 < 3^3$).

Here's a method showing that it can indeed be accomplished in three weighings. First, pick two sets of four at random and compare them. If they are of equal weight, the odd ball is in the remaining four. Now pick three of these and weigh them against three balls known to be normal. If both sets are of equal weight, the oddball is the remaining one, and weighing it against a normal ball will determine if its heavy or light. If the sets are not equal, we have learned whether the oddball is heavy or light and it can only be one of the remaining three. Weighing one of them against the other determines which is which. Going back to the case in which the two sets of four are unequal, we know that the oddball is either light and one of the four light-weighing balls or heavy and one of the heavy-weighing set. Now break the light set into two sets of two and then add a random possibly heavy ball to each group. Weigh them. If they are of equal weight, the oddball is one of the remaining two, and since we know whether each one could be light or heavy, one further measurement sufficies to determine the oddball. If the light set turns out to be the two possibly light balls and one possibly heavy, then we know that it could not have been the possibly heavy ball, nor could it have been either of the two possibly light balls in the other set. But it could have been the two possibly light balls in the light set, or the possibly heavy ball in the heavy set. Now there are three possibilities remaining, and we only need to measure the two possibly light balls to determine which is which.

Problem 5.2 Data Processing Inequality

Random variables X, Y, Z form a Markov chain $X \to Y \to Z$ if the conditional distribution of Z depends only on Y: $p(z|x,y) = p(z|y)$. The goal in this exercise is to prove the data processing inequality, $I(X:Y) \geq I(X:Z)$ for $X \to Y \to Z$.

a) First show the chain rule for mutual information: $I(X:YZ) = I(X:Z) + I(X:Y|Z)$, which holds for arbitrary X, Y, Z. The conditional mutual information is defined as

$$I(X:Y|Z) = \sum_z p(z) I(X:Y|Z=z) = \sum_z p(z) \sum_{x,y} p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)}.$$

Solution: First observe that $\frac{p(x,y|z)}{p(y|z)} = \frac{p(x,y,z)}{p(y,z)} = p(x|y,z)$, which means $I(X{:}Y|Z) = H(X|Z) - H(X|YZ)$. Then

$$I(X{:}YZ) = H(X) - H(X|YZ) = H(X) + I(X{:}Y|Z) - H(X|Z) = I(X{:}Z) + I(X{:}Y|Z).$$

b) Next show that in a Markov chain $X \to Y \to Z$, X and Z are conditionally independent given Y; that is, $p(x,z|y) = p(x|y)p(z|y)$.

Solution:

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x,y)p(z|x,y)}{p(y)} = \frac{p(x|y)p(y)p(z|y)}{p(y)} = p(x|y)p(z|y).$$

c) By expanding the mutual information $I(X:YZ)$ in two different ways, prove the data processing in equality.

Solution: There are only two ways to expand this expression:

$$I(X{:}YZ) = I(X{:}Z) + I(X{:}Y|Z) = I(X{:}Y) + I(X{:}Z|Y).$$

Since X and Z are conditionally independent given Y, $I(X{:}Z|Y) = 0$. Meanwhile, $I(X{:}Y|Z) \geq 0$, since it is a mixture (over Z) of positive quantities $I(X{:}Y|Z=z)$. Therefore $I(X{:}Y) \geq I(X{:}Z)$.

Problem 5.3 Fano's Inequality

Given random variables X and Y, how well can we predict X given Y? Fano's inequality bounds the probability of error in terms of the conditional entropy $H(X|Y)$. The goal of this exercise is to prove the inequality

$$P_{\text{error}} \geq \frac{H(X|Y)-1}{\log|X|}.$$

a) Representing the guess of X by the random variable $\widehat{X}$, which is some function, possibly random, of Y, show that $H(X|\widehat{X}) \geq H(X|Y)$.

Solution: The random variables X, Y, and $\widehat{X}$ form a Markov chain, so we can use the data processing inequality. It leads directly to $H(X|\widehat{X}) \geq H(X|Y)$.

b) Consider the indicator random variable E which is 1 if $\widehat{X} \neq X$ and zero otherwise. Using the chain rule we can express the conditional entropy $H(E,X|\widehat{X})$ in two ways:

$$H(E,X|\widehat{X}) = H(E|X,\widehat{X}) + H(X|\widehat{X}) = H(X|E,\widehat{X}) + H(E|\widehat{X}) \tag{1}$$

Calculate each of these four expressions and complete the proof of the Fano inequality. Hints: For $H(E|\widehat{X})$ use the fact that conditioning reduces entropy: $H(E|\widehat{X}) \leq H(E)$. For $H(X|E,\widehat{X})$ consider the cases $E=0,1$ individually.

Solution: $H(E|X,\widehat{X}) = 0$ since E is determined from X and $\widehat{X}$. $H(E|\widehat{X}) \leq H(E) = h_2(P_{\text{error}})$ since conditioning reduces entropy.

$$\begin{aligned} H(X|E,\widehat{X}) &= H(X|E=0,\widehat{X})p(E=0) + H(X|E=1,\widehat{X})p(E=1) \\ &= 0(1-P_{\text{error}}) + H(X|E=1,\widehat{X})P_{\text{error}} \leq P_{\text{error}} \log|X| \end{aligned}$$

Putting this together we have

$$H(X|Y) \leq H(X|\widehat{X}) \leq h_2(P_{\text{error}}) + P_{\text{error}} \log|X| \leq 1 + P_{\text{error}} \log|X|,$$

where the last inequality follows since $h_2(x) \leq 1$. Rearranging terms gives the Fano inequality.