

Lecture 12, Tuesday w7, 2014-10-28

We reviewed discrete memoryless channels and discussed all the many different conditional, marginal and joint probabilities and entropies. The definitions are all summarized at the end of the week 6 slides. If you draw the block diagram, you can read off the three expressions for mutual information. Some grunt work: you should know all of these definitions and how they fit together.

The capacity: the maximum possible mutual information for a channel, achieved by the *optimal input distribution*.

The mutual information is positive:

- Proof: compare $P(x, y)$ and the independent distribution $P(x)P(y)$ with KL. The result drops out by Gibbs inequality.
- Implication: observing data y , *on average*, cannot increase our uncertainty about any other quantity.

Check your progress

We just started the ‘week 7’ slides. Mark anything that’s unclear or that needs expanding on NB.

You can also review all of the quantities for dependent variables in Chapter 8 of MacKay. (We won’t use the three-term conditional mutual information $I(X; Y|Z)$ in this course.) There are exercises to check your understanding.