**Bayesian Probability**

* *When Bayesians talk about probability they mean something different than what is meant by probability in ordinary statistics. On the other hand, Bayesian notion of probability is probably closer to the meaning of the word probability in English language.*
* *Bayesian probability is subjective but not arbitrary*

In Bayesian probability theory we handle probabilities of statements. Probabilities tell how certain we are that those statements are true. Probability one means that we are absolutely certain. Probability zero also means that we are absolutely certain, but this time we are absolutely certain that the statement is false. Probability 0.5 means that we are maximally uncertain whether the statement is true or false.

Because probabilities are interpreted as certainties Bayesian probability is called subjective probability.

Having subjective probability makes it possible to talk about probabilities of any things we feel uncertain about. For example we can talk about the probability that there is life outside of our own solar system. In classical probability theory this kind of probability cannot be discussed, since there either is or is not, but there is nothing "random" in it. Bayesians agree, but continue that they still do not know, whether there is or not and that is why they can use probabilities. Furthermore, this probability will change, when we get more information.

Changing probabilities when getting new information is the core of ***Bayesian reasoning***. The so called Bayes rule defines how rational agent changes its believes when she/he gets new information (evidence).

Without this rule Bayesian probabilities would be "mere" subjective uncertainties and it would be hard to imagine what role that type of "arbitrary" certainties had to do with science. B-Course is initially very uncertain about the right model, but when it receives data it uses Bayes rule to update its beliefs about different models. It then picks the one it "feels" most certain about.

***Conditional Probabilities***

Bayesian probabilities, certainties, are always conditional. That means that probabilities are estimated in the context of some background assumptions. For example when estimating the probability of rain next Sunday we probably use our background knowledge about the weather patterns of the current season and so on.

Conditional probabilities of things are written with notation ***P(Thing | Assumption).*** These probabilities are actually numbers between zero and one, that tell how certain we are that ***Thing*** is true, when we believe that the assumptions are true.

We often write ***P(D | M)*** or ***P(M | D),*** where ***M*** is dependency model and ***D*** is your data. ***P(D | M)*** means the probability to get (see/have) the data ***D*** if we believe that model ***M*** is the true model. Likewise, ***P(M | D)*** means the probability that model ***M*** is the true model, when we have the data ***D***.

Sometimes we also write ***P(M)*** and ***P(D)***. These are sloppy Bayesian notations, since all the probabilities should are actually conditional, but sometimes, when all the terms have same background assumptions, we do not bother to write them down.

Actually, we should always write ***P(D | M, U)*** and ***P(M | D, U)*** and ***P(M | U)*** and ***P(D | U)***, where ***U*** is a set of background assumptions, that contains many kind of things like the assumption that true model is in the set of models, that we do not have missing data etc.

**Defining vs. calculating probabilities**

Bayesian theory is philosophically rather simple, but it can sometimes be mathematically very cumbersome, which sometimes makes people to think that Bayesian theory is hard to understand.

However Bayesian ideas can be understood without going through all that maths, like one can understand what it means to say "*number of fish in a lake*" even if it may be very difficult to find out what that actual number is.

For example something like probability of the data given the model ***P(D | M)*** can be very hard to calculate, but it is not that hard to understand that there is such thing than probability of the data given the model. For example if we have a fair coin(that is our model now) and we toss it 100 times, it is more probable to get 50 heads and 50 tails than to get 100 heads and no tails.

Clearly some data sets seem to be more probable than the others and we soon realize that different data sets may have different probabilities, even if we do not know how these different probabilities are actually calculated.