Bayes' rule simply

March 24, 2019

Bayes' rule is usually written

$$P(\theta|x) = P(x|\theta)\frac{P(\theta)}{P(x)}$$

In practice we're trying to learn about some model parameter θ given some observation x. The model $P(x|\theta)$ tells us how observations are influenced by the model parameter. This seems simple enough, but a small change in notation reveals how simple Bayes' rule is. Let us call $P(\theta)$ the prior on θ and $P'(\theta)$ the posterior on theta. Then Bayes' rule says:

$$P'(\theta) \propto P(x|\theta)P(\theta)$$

We got rid of the denominator $P(x)$ because it's just a normalisation to make the total probability sum to 1, and instead say that $P'(\theta)$ is proportional to $P(x|\theta)P(\theta)$. The value $P(x|\theta)P(\theta) = P(x, \theta)$ is the joint probability of seeing a given pair (x, θ), so we can also write Bayes' rule as:

$$P'(\theta) \propto P(x, \theta)$$

So up to normalisation, the posterior is just substituting the actual observation $X = x$ into the joint distribution. How can we interpret this? Imagine that we have a robot whose current state of belief is given by $P(x, \theta)$ and that x, θ only have a finite number of possible values, so that the robot has stored a finite number of probabilities $P(x, \theta)$, one for each pair (x, θ). Suppose that the robot now learns $X = x$ by observation. What does it do to compute its posterior belief? It first sets $P(y, \theta) = 0$ for all $y \neq x$ because the actual observed value is x. Then it renormalises the probabilities to make $P(x, \theta)$ sum to 1 again. That's all Bayes' rule is: simply delete the possibilities that are incompatible with the observation, and renormalise the remainder.