Bayesian inference

1. Bayes' Law

   a. One of the most useful formulas in basic probability theory is **Bayes' law** or **Bayes' theorem**. In particular, we will find it to be especially important for cognitive science, where it is used extensively in models of reasoning with uncertainty. Bayes' law states that:
   P(A|B) = P(B|A)P(A)/P(B)

   b. Thus, Bayes' law expresses the relationship between the conditional probability P(A|B) and the reverse conditional P(B|A). At first blush, there is nothing special about Bayes' law, as it can be seen to follow naturally from the basic definitions of conditional probability. However, as we will see, it becomes very useful when we add a certain interpretation of probability into our calculations.

2. Bayesian inference

   a. Bayes' law gets its usefulness for cognitive science when we make some assumptions about the variables we are considering and the meaning of probability. In particular:

      i. We have an agent who is trying to infer the process that is responsible for generating some data *d*.

      ii. The agent has several hypotheses (we'll consider two, *h* and *h'*) about the process that generated the data

      iii. The agent uses probability to represent the degree of belief in each of the hypotheses

      iv. P(h) is the probability that the agent ascribes to h being the true generating process, prior to (or independent of) seeing the data *d*. This quantity is known as the **prior probability** of h. Similarly, P(h') is the agent's prior degree of belief in h'.

      v. The question for Bayesian inference is this: How should the agent change his beliefs in light of the evidence provided by d?

      vi. To answer this question, we compute the **posterior probability** P(h|d), or the degree of belief in h conditioned on the observation of d.

      vii. Bayes' rule allows us to compute this by treating both the hypotheses and the data as random variables. So, to calculate the posterior probability of h, we can apply Bayes' law:

         1. P(h|d) = P(d|h)P(h)/P(d)

      viii. In this equation, the term P(d|h) is called the **likelihood**, since it represents the likelihood of seeing the data given the hypothesis.

      ix. P(d) is referred to as the **marginal probability** of the data. The marginal is calculated using the Law of Total Probability, summing up the probability of seeing the data under each hypothesis, times the probability of seeing that hypothesis. So, for two hypotheses *h* and *h'*, we get:

         1. P(d) = P(d|h)P(h) + P(d|h')P(h')

x. Thus, according to Bayes' rule, the posterior probability of a hypothesis *h* is directly proportional to the product of its prior probability times its likelihood, relative to the sum of these same scores (products of priors and likelihoods) for all alternative hypotheses under consideration.

3. Thus, under its Bayesian inference interpretation, Bayes' law describes how we should *rationally* update our beliefs in light of new data. In other words, it is *prescriptive* of how we should reason under uncertainty, in the same way that deductive logic was prescriptive of how we should reason under certain conditions.
4. Bayesian ideas have been used to model a variety of phenomena in cognitive science involving reasoning under uncertainty. For example, some of the questions that Bayesian models have been used to explore:
   a. How do people recognize words from noisy speech?
   b. How do children learn rules of grammar from finite samples of noisy speech?
   c. How do people learn categories, based only on instances of category membership?