Bayesian Tutorial Exercises

1) **Derive Bayes' rule:** Use only the sum and product rules of probability:

   - **Sum rule**: $P(A) = \sum_{B} P(A,B)$

   - **Product rule**: $P(A,B) = P(A|B)P(B) = P(B|A)P(A)$

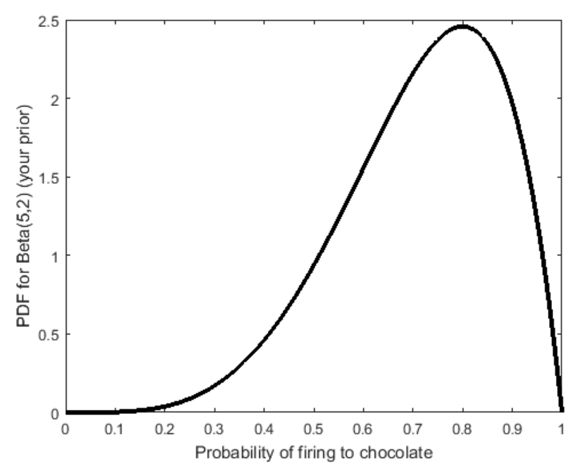
   - **Bayes' rule**: $P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in H} P(d|h')P(h')}$

2) **A Bayesian estimator:** You want to estimate some quantity S based on a noisy measurement M. Show that $E[S|M]$ is the minimum mean squared error (MMSE) estimator for S. That is, show that:
   - $argmin_{\hat{S}} E[(\hat{S} - S)^2 | M] = E(S|M)$
   - Hint: set the derivative to 0 to find where the minimum is.
   - Why would we consider this estimator to be Bayesian? (Hint: assume that you have characterized the noise model for your measurement as $P(M|S)$ but don't know $P(S|M)$ directly.

3) **Apply the Bayesian approach to estimate sound location.** You hear a sound coming from somewhere to the left of your head, and want to know how much to turn your head to shift your gaze to the source of the sound. You know that the location of the sound source varies trial by trial -- it is sampled from a (prior) Gaussian distribution with a mean of 20 deg, and a std of 8 deg. Because your sensory system is also noisy, you cannot fully rely on your auditory input. Let's assume that you know that your sensory measurement is perturbed by zero-mean Gaussian noise with a std of 3 deg.
   - *Maximum-Likelihood (ML) Estimate:* If you ignore your prior knowledge about the sound source, a common estimator is the value which maximizes the likelihood function $f(S) = P(M|S)$ (Note: $f(S)$ is <u>not</u> a probability distribution). Using maximum-likelihood estimation, where would you say the source is if you hear it come from 18.5 deg to the left? (*Bonus:* would you say the same if I told you that noise in the sensory system is higher for more peripheral sounds?)

- *Bayesian (MMSE) Estimate:* How would you change your estimate, if you additionally use your prior knowledge? Is it any different? Is it any better? Why?
- *Bonus*: Repeat the following experiment 100 times: for each trial, assume that the location of sound source is at 19 deg left and simulate a sensory measurement of the source (perturbed by sensory noise); then compute the ML and the MMSE estimates from this measurement. Compare the mean and distribution of estimates derived from the ML and the MMSE estimators. Can you now say why/how the MMSE estimate might be better?

4) **Will your neuron fire? Updating beliefs:** Consider a population of olfactory neurons that elicit a spike to various odors in a probabilistic fashion.
   - You test one of these neuron's responses to a chocolate smell and see that on 750 of 1000 trials, the neuron fires a spike. What is the probability that this neuron will fire in response to chocolate smell?
   - Now imagine a case where after the first trial, you lose the neuron, so you don't get to do multiple trials, and you are left with the impoverished information that the neuron did fire in response to chocolate smell the one time you tested it. What would you estimate is the probability of firing for this neuron?
   - To better inform your estimate, you go to your PI and ask for a record of all the neurons previously recorded in this brain in response to chocolate, and your PI gives you a whole distribution of firing probabilities across neurons. How can you use this information to get a better estimate for the probability of firing of the neuron that you only recorded one trial for?
   - Your PI quantified the full distribution of firing probabilities across neurons by a Beta distribution shown below (use this as your prior):



- Use the result of your one-trial experiment to plot the likelihood function. Now use the Bayesian approach to derive the MMSE estimate of your neuron's firing probability.

5) **Planning when your error depends on your plan:** You are playing a game of one-dimensional horseshoes, where you are penalized proportional to the square of the distance to the target (i.e. you want to minimize the mean square error). Assume that your throws can be modeled with a gaussian distribution, where the standard deviation scales with the distance that you aim for. That is, the actual distance you throw is given by $D_{actual} = D_{aim} + \varepsilon$, where $\varepsilon \sim N(0,\ cD_{aim})$. Assuming you know the correct distance, D, what D_{aim} should you aim for?