HW 1 Problems

Gamma-Exponential model

We write $X \sim \text{Exp}(\theta)$ to indicate that X has the Exponential distribution, that is, its p.d.f. is

$$p(x|\theta) = \text{Exp}(x|\theta) = \theta \exp(-\theta x) \mathbb{1}(x > 0).$$

The Exponential distribution has some special properties that make it a good model for certain applications. It has been used to model the time between events (such as neuron spikes, website hits, neutrinos captured in a detector), extreme values such as maximum daily rainfall over a period of one year, or the amount of time until a product fails (lightbulbs are a standard example).

Suppose you have data $x_1, \ldots, x_n$ which you are modeling as i.i.d. observations from an Exponential distribution, and suppose that your prior is $\boldsymbol{\theta} \sim \text{Gamma}(a, b)$, that is,

$$p(\theta) = \text{Gamma}(\theta|a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \mathbb{1}(\theta > 0).$$

1. Derive the formula for the posterior density, $p(\theta|x_{1:n})$. Give the form of the posterior in terms of one of the distributions we've considered so far (Bernoulli, Beta, Exponential, or Gamma).

2. Now, suppose you are measuring the number of seconds between lightning strikes during a storm, your prior is $\text{Gamma}(0.1, 1.0)$, and your data is

   $$(x_1, \ldots, x_8) = (20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0).$$

   Using the programming language of your choice, plot the prior and posterior p.d.f.s. (Be sure to make your plots on a scale that allows you to clearly see the important features.)

3. Give a specific example of an application where an Exponential model would be reasonable. Give an example where an Exponential model would NOT be appropriate, and explain why.

Decision theory

4. Show that if ℓ is $0-1$ loss and S is a discrete random variable, then the action a that minimizes the posterior expected loss $\rho(a, x_{1:n}) = \mathbb{E}(\ell(S, a)|x_{1:n})$ is the a that maximizes $\mathbb{P}(S = a \mid x_{1:n})$.

5. Consider the Beta-Bernoulli model. Intuitively, how would you predict x_{n+1} based on observations $x_1, \ldots, x_n$? Using your result from Question 1, what is the Bayes procedure for making this prediction when ℓ is $0-1$ loss?

6. Are there settings of the "hyperparameters" a, b for which the Bayes procedure agrees with your intuitive procedure? Qualitatively (not quantitatively), how do a and b influence the Bayes procedure?

7. What is the posterior mean $\mathbb{E}(\boldsymbol{\theta}|x_{1:n})$, in terms of a, b, and $x_1, \ldots, x_n$? Express this as a convex combination of the sample mean $\bar{x} = \frac{1}{n}\sum x_i$ and the prior mean (that is, write it as $t\bar{x} + (1-t)\mathbb{E}(\boldsymbol{\theta})$ for some $t \in [0, 1]$).