STAT 339: HOMEWORK 5 (BAYESIAN INFERENCE BASICS)

DUE VIA GITHUB SUNDAY 11/21

Instructions. Create a directory called `hw5` in your `stat339` GitHub repo. Your main writeup should be called `hw5.pdf`.

You may also use any typesetting software to prepare your writeup, but the final document should be a PDF. LaTeXis highly encouraged.

I will access your work by cloning your repository; make sure that any file path information is written relative to your repo – don't use absolute paths on your machine, or the code won't run for me!

1. **Bayesian inference for a proportion.** In order to determine how effective a magazine is at reaching its target audience, a market research company selects a random sample of N people from the target audience and interviews them. Let μ represent the **proportion of the target audience that has seen the latest issue** and Y be the random variable representing **number in the interview group who has seen it**.

   (a) Since the respondents are modeled as a random sample from the audience, we can model the conditional distribution of Y given μ using a **Binomial distribution**, with N independent trials, each having "success chance" μ. We have previously seen that the MLE for μ would be $\frac{y}{N}$, where y is the observed value of Y. Let's take a Bayesian approach here and compare it to the MLE estimate.

   Using a **continuous uniform** prior on μ (on the interval $[0, 1]$), find the **posterior density** of μ in terms of y and N.

   (b) Find the posterior mean of μ, $\mathbb{E}[\mu|Y = y]$, in terms of y and N. (Hint: The posterior density is a member of a named family of distributions. Refer to HW3 to find its mean – you don't have to re-derive it!)

Date: Last Revised: November 15, 2021.

(c) Represent $\mathbb{E}[\mu|Y=y]$ as a weighted average of two terms: the **MLE** for μ, $\hat{\mu}$, and the **prior mean**, $\mathbb{E}[\mu]$. That is, find an expression for $\alpha \in [0,1]$ so that $\mathbb{E}[\mu|Y=y] = \alpha\mathbb{E}[\mu] + (1-\alpha)\hat{\mu}$.

(d) Show that the **prior density** in 1a (which was a continuous uniform on $[0,1]$) is a special case of a **Beta distribution**. Find its parameters, and generalize the result in 1e for an arbitrary choice of Beta distribution prior with parameters a and b.

(e) What does the expression you found for the weight α in suggest about the interpretation of the parameters of the Beta prior? **Hint:** think about how incrementing or decrementing these parameters would "trade off" with incrementing or decrementing y or N.

2. **A waiting time model.** The **exponential distribution** with **rate parameter** λ is a density for a **continuous random** variable whose range is **all nonnegative real numbers**. This distribution often used to model the **amount of time that passes between two events**. Its PDF is

$$p(y|\lambda) = \lambda e^{-\lambda y}$$

(a) This density is a special case of a $\mathsf{Gamma}(a,b)$ density. A random variable y has a **Gamma distribution** with parameters a and b if its density is

$$p(y \mid a,b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by} I(y>0)$$

where $\Gamma()$ is the **gamma function** (see HW3), and $I(A)$ is the **indicator function** which is equal to 1 if the proposition A is true and 0 otherwise. Here, this restricts the density to be nonzero only on the nonnegative reals.

Find expressions for the values of a and b written in terms of λ such that the Gamma density is equivalent to an Exponential density with rate λ.

(b) Show that the $\mathsf{Gamma}(a,b)$ family is also a **conjugate prior** for the rate parameter of the Exponential. That is, if the prior density on λ has the form

$$p(\lambda; a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} \cdot \lambda^{a_0-1} e^{-b_0\lambda}$$

where $a_0, b_0 > 0$ are prior parameters, then the posterior density $p(\lambda|y)$ has a $\Gamma(a_1, b_1)$ density. Find expressions for a_1 and b_1 in terms of y, a_0 and b_0 (**Hint:** Remember that the non-constant part of a density function determines the density function: the value of any constant factor is constrained by the fact that the density must integrate to 1)

(c) Suppose $Y_1, \ldots, Y_N$ are independent and identically distributed random variables, each with the same **Exponential density** with rate parameter λ. Show that if the **prior density** on λ, is $\mathsf{Gamma}(a_0, b_0)$, then the **posterior density**, $p(\lambda|y_1, \ldots, y_N)$ is $\mathsf{Gamma}(a_N, b_N)$, and find expressions for a_N and b_N in terms of a_0, b_0, $y_1, \ldots, y_N$, and N.

(**Hint:** First write down the joint density for $Y_1, \ldots, Y_N$ conditioned on λ, which defines the likelihood function)

3. **Inferring a Detection Limit.** Suppose the random variable Z represents the **amount of radiation** in an area, which is modeled with an Exponential distribution with rate parameter λ. That is, the density of Z is

$$p(z) = \lambda e^{-\lambda z}.$$

Suppose also that an instrument only registers the presence of radiation if it is above a threshold θ. That is, if the underlying quantity is below θ, no observation is produced, so that when a measurement *is* taken, the reported amount is always at least θ.

Let Y be the **measured amount of radiation**.

(a) Find a formula for the **conditional density** of Y given that radiation is reported; that is, conditioned on the event $Z > \theta$.

(b) Suppose λ is known, but θ is not. **Find a formula for the likelihood function**, $L(\theta; y) := p(y \mid \theta)$, given a single observation $Y = y$. Find a formula for the MLE, $\hat{\theta}$ (Don't forget that the range of Y is restricted!)

(c) Suppose we want to use Bayesian inference to estimate the minimum detectable quantity, θ. The **prior support** of θ, that is, the range in which the prior density is positive, is $[0, \infty)$. Given an observation, $Y = y$, what is the **posterior support** for θ? (We have not specified a prior density for θ yet; we do not need to know the actual density to answer this, only that it is positive for all $\theta \in [0, \infty)$)

(d) Suppose θ has a Gamma prior: $\theta \sim \mathsf{Gamma}(a_0, b_0)$, with prior density

$$p(\theta) = \left\{ \frac{b_0^{a_0}}{\Gamma(a_0)} \theta^{a_0 - 1} e^{-b_0 \theta} I(\theta > 0) \right.$$

Find the **posterior PDF**, $p(\theta \mid y)$, up to a normalization constant, $k(a_0, b_0, y, \lambda)$. Was the prior conjugate to the likelihood? That is, is the posterior also a Gamma distribution (for some values of a and b)? How do you know?

4. **Posterior for a Poisson parameter.** Consider the fish taco model in which we assume that the number of customers that buy fish tacos in a given hour can be modeled by a Poisson distribution with parameter λ. Let $Y_1, \ldots, Y_N$ be the number of tacos sold in hours 1 through N. The conditional PMF of $Y_n \mid \lambda$ is then

$$p(y_n \mid \lambda) = \frac{e^{-\lambda}\lambda^{y_n}}{y_n!}$$

   (a) Assuming $Y_1, \ldots, Y_N$ are independent and identically distributed, show that the likelihood function $L(\lambda; y_1, \ldots, y_N)$ is equal to $C \cdot L(\lambda; s)$, if $S = \sum_{n=1}^{N} Y_n$, the total number of fish tacos sold over the N hours, is modeled using a Poisson distribution with parameter $\lambda^* := N\lambda$, and C is a constant that does not depend on λ (but can depend on $Y_1, \ldots, Y_N$)

   (b) The conjugate prior for the Poisson parameter, λ is also a Gamma distribution. Suppose the prior on λ is $\mathsf{Gamma}(a_0, b_0)$; that is

$$p(\lambda) = \frac{b_0^{a_0}}{\Gamma(a_0)}\lambda^{a_0-1}e^{-b_0\lambda}$$

   Find the posterior parameters in terms of the prior parameters a_0 and b_0 and the data values, $\{y_1, \ldots, y_N\}$.

5. **A naive and inefficient method to approximate the posterior.** In the last problem we were able to find the posterior distribution analytically. But for many models this will not be true, and we will often need to resort to approximation methods to do computations with the posterior distribution. An extremely naive (and inefficient) method is to take many samples from the **joint distribution** of the parameters and data (that is, generate a sequence of pairs $\{(\lambda_t^{(sim)}, s_t^{(sim)})\}$, $t = 1, \ldots, T$), and "condition" by retaining only those samples that yield data values identical to those observed (that is, where $s_t^{(sim)} = s$, where s is the actual number of tacos sold). The posterior distribution is then approximated by the collection of $\lambda_t^{(sim)}$ values from the retained pairs (that is, the set of $\lambda_t^{(sim)}$ such that $s_t^{(sim)} = s$).

   (a) Implement this method using the following algorithm, which takes as inputs the values s, a_0, b_0, N, T_{target} and T_{max} and returns an array of $\lambda_t^{(sim)}$ values.

      - Initialize $t_{kept} = 0$
      - While $t \leq T_{max}$ and $t_{kept} < T_{target}$:

(i) Sample a value of $\lambda_t^{(sim)}$ from the prior $\mathsf{Gamma}(a_0, b_0)$ distribution (use `numpy.random.gamma()` to generate Gamma-distributed values – note that this function has arguments `shape` and `scale`, which correspond to a_0 and $1/b_0$, respectively. That means you'll need to take a reciprocal of b_0 before calling it)

(ii) Using the sampled value of λ, sample a value of $S_t^{(sim)}$ from the conditional distribution of S given λ, which is a $\mathsf{Poisson}(N\lambda)$ distribution (use `np.random.poisson()`).

(iii) If $S_t^{(sim)} = s$, add $\lambda_t^{(sim)}$ to the kept values and increment t_{kept}. Otherwise, discard $\lambda_t^{(sim)}$ and do not increment t_{kept}.

(iv) Increment t.

(b) Run your algorithm setting $s = 10$, $N = 5$, $T_{target} = 1000$ and $T_{max} = 100000$, and letting a_0 and b_0 take the pairs $(2, 0.1)$, $(4, 0.2)$ and $(100, 50)$. Plot both the theoretical posterior density and a histogram of the simulated posterior samples.

```
import runpy
p6_globals = runpy.run_path("./problem6.py")
```

(c) Compare the theoretical and simulated means. Are they close?