STAT 339: HOMEWORK 3 (BAYESIAN INFERENCE AND BELIEF NETWORKS)

DUE VIA GITHUB AROUND WEDNESDAY 4/8

Instructions. Create a directory called `hw3` in your `stat339` GitHub repo. Your main writeup should be called `hw3.pdf`.

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 number in the interview group who has seen it.
   (a) Using a uniform prior on μ, find the posterior distribution of μ in terms of Y and N.
   (b) Find $\mathbb{E}[\mu|Y=y]$ in terms of y and n.
   (c) Rewrite $\mathbb{E}[\mu|Y]$ as a weighted average of two terms: the MLE $\hat{\mu}$, and the prior mean, $\mathbb{E}[\mu]$. That is, find an expression for α so that $\mathbb{E}[\mu|Y=y] = \alpha\mathbb{E}[\mu] + (1-\alpha)\hat{\mu}$.
   (d) Show that the prior in 1a is a special case of a Beta distribution (see HW2a, Problem 7), and generalize the result in 1c for arbitrary choices of prior parameters.
   (e) What does the expression for α suggest about the interpretation of the prior parameters (specifically, the way they affect posterior inferences about μ)?
2. **Inferring a Detection Limit.** Suppose Y represents a non-negative measurement (i.e, $Y \geq 0$) that is that is detectable only up to a threshold θ. Suppose also that Y is uniformly distributed on its range, i.e., $Y \sim \mathsf{Unif}(0,\theta)$, but that the value of θ is unknown.
   (a) Find a formula for the likelihood function for θ, given a single observation Y. Be careful to specify the domain of the function! Where is it largest? (You

Date: Last Revised: April 8, 2020.

do not need calculus to answer this — draw the graph if you aren't sure.) Explain why, intuitively, it is largest there.

(b) Suppose we want to use Bayesian inference to estimate the upper bound, θ. The prior range of θ is $[0, \infty)$. Given an observation, $Y = y$, what is the posterior range for θ?

(c) Suppose θ has a Gamma prior: $\theta \sim \mathsf{Gamma}(a, b)$, with prior density

$$p(\theta) = \begin{cases} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} & \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the non-constant component of the posterior PDF. Is the posterior also a Gamma distribution (for some values of a and b)? How do you know?

3. **A waiting time model.** The *exponential distribution* with *rate parameter* λ and range $[0, \infty)$ is often used to model the amount of time that passes between two events. Its density is given by

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

(a) This density is a special case of a $\mathsf{Gamma}(a, b)$ density. Find the values of a and b that make the densities equivalent.

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

$$p(\lambda) = k_{a,b} \cdot \lambda^{a-1} e^{-b\lambda}$$

where $a, b > 0$ are parameters and $k_{a,b} > 0$ is a normalizing constant that depends on a and b but not on λ, then the posterior density $p(\lambda|y)$ has the same form, for different values of a, b and k.

(c) Show that if $Y_1, \ldots, Y_n$ are i.i.d. exponential with rate parameter λ, then using a $\mathsf{Gamma}(a, b)$ prior for λ results in a Gamma posterior density, $p(\lambda|y_1, \ldots, y_n)$, and find its parameters. (First write down the joint density for $Y_1, \ldots, Y_n$ / likelihood function for λ.)

4. **Simulating the Posterior for a Poisson parameter.** Consider the fish taco model from HW2a, problem 3. 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, which has a PMF given by

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

(where $y!$ is y-factorial: $y! := y \times (y-1) \times \cdots \times 2 \times 1$) We now want to use data to infer λ.

(a) Suppose we have a dataset with customer counts for N different hours, represented by $\{Y_1, \ldots, Y_N\}$. Assuming the observations are independent given λ, and that λ does not change, find the likelihood function for λ.

(b) Find the MLE, $\hat{\lambda}$ (Hint: As always, the λ that maximizes the log likelihood also maximizes the likelihood).

5. An alternative way to model arrival numbers to the fish taco stand in the previous problem is to consider the total number of customers in the entire N hour period as a single data point, generated from a Poisson distribution with mean N times as large as the single-hour distribution (i.e., $N\lambda$).

(a) Show that the likelihood function for λ in this scenario is a constant multiple of the likelihood function in the original scenario (and hence both MLE and Bayesian inferences about λ (for any prior) will be identical in both cases).

(b) Comment on what this means about the information contained in the order of the observations as it concerns inferences about λ in this model.

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

$$p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

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

6. In the last problem, since we are using a conjugate prior, we can find the posterior distribution analytically. But for many less simplistic models this will not be true. Hence we will often 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 $\{(\lambda_t^{(sim)}, y_t^{(sim)})\}$, $t = 1, \ldots, T$), and "condition" by retaining only those samples that yield data values identical to those observed (that is, where $y_t^{(sim)} = y_N$, where y_N is the real data). The posterior distribution is then approximated by the distribution of the parameter values used for the subset of the simulated data that is retained (that is, the set of $\lambda_t^{(sim)}$ such that $y_t^{(sim)} = y_N$).

(a) Implement this method using the formulation in 5a in which the data consists of a single count assumed to be drawn from a $\mathsf{Poisson}(N\lambda)$ distribution. Your function should take as input the prior parameters, a and b, the data value y_N, and a "stopping count", T_{kept}, which governs how many "retained" $(\lambda_t^{(sim)}, y_t^{(sim)})$ pairs must be generated before the algorithm stops.

(b) Run your algorithm for a few different combinations of y_N, a and b values, plotting both the theoretical posterior density and a histogram of the simulated posterior samples.

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

7. Consider the Bayes net depicted in Fig. 1, which comes from the BRML book. Each variable is binary.

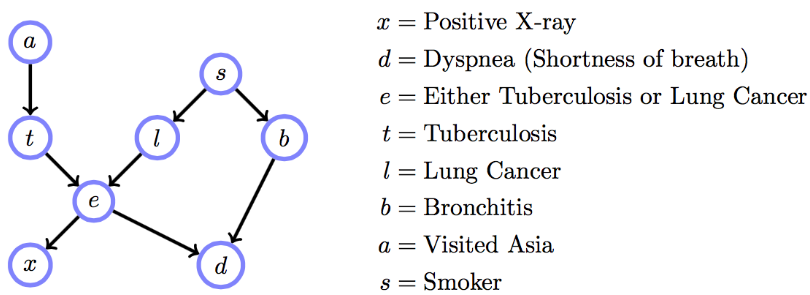

Figure 3.15: Belief network structure for the Chest Clinic example.

FIGURE 1. Bayes Net for diagnosis of lung disease at a chest clinic

(a) Write down the factorization of the joint distribution that is implied by the graph.

(b) According to the model, can you predict whether someone has visited Asia based on whether or not they are a smoker? That is, are s and a independent?

(c) Does knowing that someone is a smoker help you predict whether they visited Asia if you also have a chest x-ray? That is, are s and a conditionally independent given x? Explain the intuition behind these two results.