STAT 447 Assignment 5

Caden Hewlett

2024-02-08

Question 1: Sequential Updating

Consider a joint a joint probabilistic model given by

$$\theta \sim \rho, \text{ and } (x_i \mid \theta) \overset{iid}{\sim} \nu_\theta, \text{ where } i \in \{1, 2, \dots, n\}$$

where ρ is a prior distribution for the unknown parameter θ, and $\{x_i\}_{i=1}^n$ is a sequence of observations with conditional distribution ν_θ.

Part 1

Write down the posterior distribution of θ given $\{x_i\}_{i=1}^n$.

In a more verbose sense, let Θ be the random variable for the unknown parameter, and θ be the realization of this random variable under the proposed prior distribution. We can then write the following (*purely for nomenclature reasons*)

$$\rho = p_\Theta(\theta), \text{ where } p_\Theta(\theta) \text{ is the PMF/PDF given by } \rho$$
$$P(X = x \mid \theta) = p_{X|\Theta}(x, \theta), \text{ where } p_{X|\Theta}(x, \theta) \text{ is the PMF given by } \nu_\theta$$

With this in mind, we can write the posterior of θ given $\{x_i\}_{i=1}^n$, where we describe the event that $\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n$ using Bayes' Rule.

We will begin by using the most verbose notation possible, for complete clarity.

$$P(\Theta = \theta \mid \{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n) = \frac{p_\Theta(\theta) P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n \mid \Theta = \theta\big)}{P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n\big)}$$

We start by considering the joint likelihood function.

$$P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n \mid \Theta = \theta\big) = P\big((X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1) \mid \Theta = \theta\,\big)$$

Then, using intersections, we can write the likelihood as:

$$P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n \mid \Theta = \theta\big) = P\Big(\bigcap_{i=1}^n (X_i = x_i) \mid \Theta = \theta\Big)$$

Now, we use the following property of *iid* random variables

I.I.D $\implies \forall (i \neq j) \in [1, n],\ (X_i \mid \Theta) \perp (X_j \mid \Theta) \implies \forall (i \neq j) \in [1, n], P(X_i \cap X_j \mid \Theta) = P(X_i \mid \Theta) P(X_j \mid \Theta)$

Hence, we can write the likelihood as:

$$P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n \mid \Theta = \theta\big) = \prod_{i=1}^{n} \Big(P(X_i = x_i \mid \Theta = \theta)\Big) = \prod_{i=1}^{n} (\nu_\theta) = (\nu_\theta)^n$$

Which, as you can see, simplifies nicely to $(\nu_\theta)^n$.

Now, our expression simplifies a little bit to the following:

$$P(\Theta = \theta \mid \{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n) = \frac{p_\Theta(\theta)\nu_\theta^n}{P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n\big)}$$

If we wished to, we could write the following proportionality directly to conclude:

$$p_{\Theta|X_{1:n}}(\theta, \{x_i\}_{i=1}^n) = \pi_n \propto \rho \cdot \nu_\theta^n$$

Or, letting normalizing constant $\mathcal{Z}_{1:n} = P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n\big)$, we can write:

$$\pi_n(\theta) = \frac{\rho \cdot \nu_\theta^n}{P\big(\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n\big)} = (\mathcal{Z}_{1:n}^{-1}) \cdot \rho \cdot \nu_\theta^n$$

Giving us a nice expression for the posterior, both in proportionality and equality according to a normalizing constant.

Part 2

Suppose now we get an additional data point x_{n+1} with the same conditional distribution ν_θ. Show that using the posterior from part 1 as the *prior* and data equal to just x_{n+1} gives the same posterior distribution as redoing part 1 with the $n+1$ data points.

We wish to show that:

$$\pi_{(n+1)}(\theta) = \frac{p_\Theta(\theta)P\big(\{X_i\}_{i=1}^{n+1} = \{x_i\}_{i=1}^{n+1} \mid \Theta = \theta\big)}{P\big(\{X_i\}_{i=1}^{n+1} = \{x_i\}_{i=1}^{n+1}\big)} = \frac{\pi_n(\theta)P(X_{n+1} = x_{n+1} \mid \Theta = \theta)}{P(X_{n+1} = x_{n+1})}$$

We'll evaluate each expression in turn.

With the first term, we can say directly that:

$$\text{LHS} = \frac{p_\Theta(\theta)P\big(\{X_i\}_{i=1}^{n+1} = \{x_i\}_{i=1}^{n+1} \mid \Theta = \theta\big)}{P\big(\{X_i\}_{i=1}^{n+1} = \{x_i\}_{i=1}^{n+1}\big)} = (\mathcal{Z}_{1:(n+1)}^{-1}) \cdot \rho \cdot \nu_\theta^{n+1} \propto \rho \cdot \nu_\theta^{n+1}$$

By the exact process used in **Part 1**.

More interestingly, we can expand the second term to as follows:

$$\pi_{(n+1)}(\theta) = \frac{\pi_n(\theta)P(X_{n+1} = x_{n+1} \mid \Theta = \theta)}{P(X_{n+1} = x_{n+1})} = (\mathcal{Z}_{n+1}^{-1})\pi_n(\theta)\nu_\theta$$

Then, substituting our expression for $\pi_n(\theta)$ from **Part 1**:

$$\text{RHS} = (\mathcal{Z}_{n+1}^{-1})\Big((\mathcal{Z}_{1:n}^{-1}) \cdot \rho \cdot \nu_\theta^n\Big)\nu_\theta = \big(\mathcal{Z}_{n+1}^{-1} \cdot \mathcal{Z}_{1:n}^{-1}\big)\rho\big(\nu^n \cdot \nu^1\big) = (\mathcal{Z}_{n+1}^{-1} \cdot \mathcal{Z}_{1:n}^{-1})^{-1}\rho \cdot \nu_\theta^{n+1} \propto \rho \cdot \nu_\theta^{n+1}$$

Question 2: Baesian Inference in the Limit of Increasing Data