Homework #3 — MTH 602

Assigned: Wednesday, October 1, 2025
Due: Monday, October 13, 2025 (print off and submit your report in class if possible)

Hardcopy reports should be submitted to Scott by Monday, October 13 at 5pm. If you cannot make it to class, you must find some other way to submit the report. **You can work in groups of 1, 2, or 3.**

Report: Put all the figures and answers asked for in the questions into a single, well-organized PDF document (prepared using LaTeX, a Jupyter notebook, or something else). **Please submit your report by uploading it to your git project and printing it off to hand in.** Ensure your submission is *reproducible*: I should be able to clone your repository and generate your reported results by using available code. Paper and pencil questions can be included in the PDF report or turned in separately

I. BAYESIAN PROBABILITY AND INFORMATION THEORY (CHAPTER 2)

Solve the following book problems (10 points each)

1. Problem 2.21
2. Problem 2.25
3. Problem 2.34
4. ~~Problem 2.40~~ [Scott: the books question is worded confusingly. If you've already done it thats OK. Otherwise, please skip this one]
5. Problem 2.41

II. BAYESIAN ESTIMATION OF A SATELLITE'S ORBITAL FREQUENCY

This toy model is a simplified version of how Bayesian techniques can be used in real navigation systems, object tracking, robotics, and similar dynamical motion problems that have a physics-based model. This application of Bayes' rule to dynamical systems is sometimes called Bayesian filtering (related to Kalman filters). The setup here is often used to forecast future positions of an object, but here we focus on extracting the orbital frequency.

Problem statement: Imagine a satellite in a circular orbit around Earth, moving in the x–y plane. If the orbital angular frequency is ω, then its position at time t is

$$x(t) = r\cos(\theta(t)), \quad y(t) = r\sin(\theta(t)),$$

with the angular coordinate

$$\theta(t) = \omega t.$$

Here we assume $\theta(t)$ (unit radians) are unwrapped (i.e., measured continuously, not reduced mod 2π). The variable r is the radius from the Earth's center, which we will take to be $r = 1$ for simplicity.

In practice, satellite positions are measured with error. Suppose instead of perfect positions, we observe noisy angular coordinates:

$$\theta_i = \omega t_i + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

at times t_i (in seconds). Here, the variables of the problem are

- ω (radians/s) is the unknown orbital angular frequency
- σ (radians) is the observational noise standard deviation (which you will estimate below)

The angular coordinate dataset is provided as a CSV file (hw3.csv in the class github project).

Background on Bayesian filtering: In this homework you are asked to use both Bayesian inference and sequential Bayesian updating. These are direct applications of topics covered in class. Let's briefly motivate them.

Suppose you are given the full observation dataset, $\{t_i, \theta_i\}_{i=1}^N$ with N observations. Bayes' rule says:

$$p(\omega \mid \{\theta_i\}_{i=1}^N) \ \propto \ p(\{\theta_i\}_{i=1}^N \mid \omega)\, p(\omega).$$

This uses *all the data at once*.

Bayesian filtering is application of Bayes' rule to the first $k < N$ observations. This is a natural setup because data often arrive over time (e.g., satellites continuously tracked by ground stations). Rather than reprocess the full dataset each time, we can update the posterior distribution incrementally:

$$\begin{aligned}
&\text{First Observation}: \quad p(\omega \mid \{\theta_i\}_{i=1}^1) \ \propto \ p(\theta_1 \mid \omega)\, p(\omega) \\
&\text{Second Observation}: \quad p(\omega \mid \{\theta_i\}_{i=1}^2) \ \propto \ p(\theta_2 \mid \omega)\, p(\omega \mid \{\theta_i\}_{i=1}^1) \\
&\text{Third Observation}: \quad p(\omega \mid \{\theta_i\}_{i=1}^3) \ \propto \ p(\theta_3 \mid \omega)\, p(\omega \mid \{\theta_i\}_{i=1}^2) \\
&\text{...and so on...}
\end{aligned}$$

The general update rule when one new observation, θ_k, arrives is

$$p(\omega \mid \{\theta_i\}_{i=1}^k) \ \propto \ p(\theta_k \mid \omega)\, p(\omega \mid \{\theta_i\}_{i=1}^{k-1}).$$

Each new observation updates our prior belief, yielding the posterior after k points. This is the essence of Bayesian filtering. The main benefits of Bayesian filtering include: you don't need to compute the entire likelihood from scratch (fast), you can update the model as data streams in, and you can monitor the uncertainty in your estimate in ω as more data arrives. This is exactly the same as our coin flipping example from class – the probability for the coin bias factor H was updated as more coin flip data was observed. You should find the posterior variance of ω keeps shrinking as more data is acquired – again similar to the coin flipping example from class.

Your job: Estimate ω from data, compare maximum likelihood and Bayesian approaches, and explore how uncertainty decreases as more observations arrive. To do this, please complete the following steps (each worth 10 points):

1. **Maximum likelihood estimate (MLE)**
   (a) Write down the likelihood function, $p(\{\theta_i\}_{i=1}^N \mid \omega)$, for the full dataset $\{\theta_i\}_{i=1}^N$.
   (b) Show that the the maximum likelihood solution is

   $$\hat{\omega}_{\text{MLE}} \ = \ \frac{\sum_{i=1}^N t_i \theta_i}{\sum_{i=1}^N t_i^2},$$

   (c) Numerically compute $\hat{\omega}_{\text{MLE}}$ from the provided dataset
   (d) Plot the noisy observations alongside the line $\hat{\omega}_{\text{MLE}} t$.
   (e) Estimate the noise variance σ^2 by $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^N (\theta_i - \hat{\omega}_{\text{MLE}} t_i)^2$. Be mindful that this assumes i.i.d. Gaussian measurement noise (which I am using). In practical applications, however, outliers, correlated noise, or model mis-specification can bias this estimate

2. **Bayesian posterior** (full dataset) Assume a Gaussian prior on ω:

   $$\omega \sim \mathcal{N}(\mu_0, \tau_0^2).$$

   Now after observing all N data points, it turns out that the posterior distribution on ω is also Gaussian (its not obvious that this simplification would happen – when it does, it's called a *conjugate prior*):

   $$p(\omega \mid \{\theta_i\}_{i=1}^N) \ = \ \mathcal{N}(\mu_N, \tau_N^2),$$

   fully specified by its variance τ_N^2

   $$\tau_N^2 \ = \ \left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\sum_{i=1}^N t_i^2\right)^{-1},$$

and by its mean μ_N

$$\mu_N \;=\; \tau_N^2 \left(\frac{\mu_0}{\tau_0^2} + \frac{1}{\sigma^2} \sum_{i=1}^{N} t_i \theta_i \right).$$

(a) Select reasonable values for the prior's parameters μ_0 and τ_0^2. Make sure your settings give a very wide prior. Use your best judgement here and justify your choice.

(b) Numerically compute the Bayesian posterior mean and variance. Print its value.

(c) Compare the Bayesian posterior mean to the MLE estimate

3. **Bayesian filtering**

(a) Write code to update the posterior one observation at a time. That is, after each new observation θ_k, compute the (i) posterior mean μ_k and variance τ_k^2. A simple approach is to maintain variables like ($S_{tt}^{(k)} = \sum_{i=1}^{k} t_i^2$) and $S_{t\theta}^{(k)} = \sum_{i=1}^{k} t_i\theta_i$, and update (μ_k, τ_k^2) from those. (Note: in the case of conjugate priors − which we have here − the benefits of sequential updating are reduced because computing the summations defining μ_k and τ_k^2 from scratch is already fast.)

(b) Plot μ_k vs k with a 95% credible interval band $\mu_k \pm 1.96\tau_k$. Please observe how quickly the uncertainty shrinks as k grows. (Note: you may need to fiddle with the y axis limits to more clearly show how the variance is changing)

4. **Information gain over time**

The entropy of a 1D Gaussian is

$$H(\mathcal{N}(\mu, \tau^2)) = \tfrac{1}{2} + \tfrac{1}{2} \ln\left(2\pi\tau^2\right).$$

(a) Set the prior's parameters to $\mu_0 = 0$ and $\tau_0^2 = 1$. Numerically compute the posterior's entropy H_k after each update.

(b) Plot entropy vs. observation number

(c) Plot entropy change ($\Delta H_k = H_k - H_{k-1}$) vs. observation number

(d) Why do earlier data reduce entropy more steeply?

III. BAYESIAN ESTIMATION OF A NEWLY DISCOVERED STAR'S MASS (10 POINTS)

A new star was discovered to have an orbiting planet at a distance $r = 1.56479e + 11$ meters from its sun! The planet is in a circular orbit in the x-y plane and obeys' Kepler's law:

$$\omega^2 = \frac{GM}{r^3}$$

Angular coordinates are measured and have the same setup as above (but σ^2 may take on a different value). An observation was recorded about once per month. The dataset is named SolarData.csv.

Use the dataset to infer a star's mass from from noisy angle measurements. **You can use any method you want.** But if you directly use the previous setup (Bayesian Posterior with the full dataset) you can (i) compute the posterior and (ii) map this posterior to mass using Kepler's law for a circular orbit at radius r by sampling $\omega \sim \mathcal{N}(\mu_N, \tau_N^2)$ and pushing those samples through $M(\omega)$.

Whats the value of the star's mass? Whats your value's error bar?

IV. CHALLENGE (+ 10 POINTS)

The group with the closest value to the true M.