# MTH 602 Scientific Machine Learning

Homework 3

10/15/2025

S. M. Mahfuzul Hasan

02181922

UMass | Dartmouth

## I. BAYESIAN PROBABILITY AND INFORMATION THEORY (CHAPTER 2)

Solution of book problems:

1. **<u>Problem 2.21</u>**

Given,
For independent variables $x$ and $y$,

$$p(x, y) = p(x)p(y) \tag{1}$$

$$h(x, y) = h(x) + h(y) \tag{2}$$

To show,

$$\begin{aligned} &h(p^2) = 2h(p) \\ &h(p^n) = nh(p) \\ &h(p^{\frac{n}{m}}) = \frac{n}{m}h(p) \\ &h(p^x) = xh(p) \\ &h(p) \propto \ln p \end{aligned}$$

If the probablity of two independent events are both $p$, then the joint probability of these two events is $p \cdot p = p^2$.
Then, using equation (2)

$$h\left(p^2\right) = h(p) + h(p) = 2h(p) \quad \text{(Showed)}$$

So, if there were $n$ such independent events with the same probability of $p$, then we can write

$$h\left(p^n\right) = h(p) + h(p) + \dots\dots + h(p) = nh(p) \quad \text{(Showed)} \tag{3}$$

Now,

$$h\left(p^{\frac{n}{m}}\right) = nh\left(p^{\frac{1}{m}}\right) = \frac{n}{m}mh\left(p^{\frac{1}{m}}\right) = \frac{n}{m}\left(p^{\frac{m}{m}}\right) = \frac{n}{m}(p) \quad \text{[using equation (3)]} \quad \text{(Showed)} \tag{4}$$

So, by continuity from equation (4), it can be shown that $h\left(p^x\right) = xh(p)$ for any positive rational number $x$.
Now, by assuming $p = q^x$ where, $p$, $q$ and $x$ are real positive numbers, we can write

$$\begin{aligned} &h(p) = h\left(q^x\right) = xh(q) = \frac{\ln p}{\ln q}h(q) \quad [\text{since } \ln p = x \ln q] \\ &\Rightarrow \frac{h(p)}{\ln p} = \frac{h(q)}{\ln q} = C, \ \text{ where } C \text{ is a constant} \\ &\Rightarrow h(p) \propto \ln p \quad \text{(Showed)} \end{aligned}$$

**2. Problem 2.25**

Given,

$$H[x] = -\int_{-\infty}^{\infty} p(x)\ln p(x)dx \tag{5}$$

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{6}$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \sigma^2 \tag{7}$$

$$p(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{8}$$

To show,

$$H[x] = \frac{1}{2}\left\{1 + \ln\left(2\pi\sigma^2\right)\right\}$$

Using equation (8) to (5),

$$\begin{aligned}
H[x] &= -\int_{-\infty}^{\infty} p(x)\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\right)dx \\
&= -\int_{-\infty}^{\infty} p(x)\left(\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\right) - \frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&= -\int_{-\infty}^{\infty} p(x)\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\right)dx + \int_{-\infty}^{\infty} p(x)\frac{(x-\mu)^2}{2\sigma^2}dx \\
&= -\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\right)\int_{-\infty}^{\infty} p(x)dx + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty}(x-\mu)^2 p(x)dx \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{[using equation (6) and (7)]} \\
&= -\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\right)\cdot 1 + \frac{1}{2\sigma^2}\cdot\sigma^2 \\
&= \frac{1}{2} - \ln 1 + \frac{1}{2}\ln\left(2\pi\sigma^2\right) = \frac{1}{2}\left\{1 + \ln\left(2\pi\sigma^2\right)\right\} \qquad \text{(Showed)}
\end{aligned} \tag{9}$$

3. **<u>Problem 2.34</u>**

Kullback–Leibler divergence is given by,

$$\begin{aligned}
\mathrm{KL}\,(p||q) &= -\int p(x) \ln\left\{\frac{q(x|\theta)}{p(x)}\right\} dx \\
&= -\int p(x) \ln\{q(x|\theta)\}\, dx + \int p(x) \ln\{p(x)\}\, dx \\
&= -\int p(x) \ln\{q(x|\theta)\}\, dx - H[x] \qquad \text{[using equation (5)]} \\
&= -\int p(x) \ln\{q(x|\theta)\}\, dx + C,
\end{aligned}$$

where $C$ is a constant as $H[x]$ is independent of $\theta$

Now, the Kullback–Leibler divergence between the empirical distribution $p(x|\mathcal{D}) = \frac{1}{N}\sum\limits_{n=1}^{N}\delta(x-x_n)$ and model distribution $q(x|\theta)$ takes the form

$$\begin{aligned}
\mathrm{KL}\,(p||q) &= -\int \frac{1}{N}\sum_{n=1}^{N}\delta(x-x_n) \ln\{q(x|\theta)\}\, dx + C \\
&= -\frac{1}{N}\sum_{n=1}^{N}\int \delta(x-x_n) \ln\{q(x|\theta)\}\, dx + C \\
&= -\frac{1}{N}\sum_{n=1}^{N} \ln\{q(x_n|\theta)\} + C \qquad \text{[using sifting property of Dirac delta function]}
\end{aligned}$$

This is the negative log likelihood function up to an additive constant. (Showed)

4. **<u>Problem 2.41</u>**

Given,

$$-\ln p(\mathbf{w}|\mathcal{D}) = -\ln p(\mathcal{D}|\mathbf{w}) - \ln p(\mathbf{w}) + \ln p(\mathcal{D}) \tag{10}$$

$$p(\mathbf{w}|s) = \prod_{i=0}^{M}\left(\frac{1}{2\pi s^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{w_i^2}{2s^2}\right\} \tag{11}$$

From equation (11),

$$\begin{aligned}
\ln p(\mathbf{w}|s) &= \ln\left[\prod_{i=0}^{M}\left(\frac{1}{2\pi s^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{w_i^2}{2s^2}\right\}\right] \\
&= \sum_{i=0}^{M} \ln\left[\left(\frac{1}{2\pi s^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{w_i^2}{2s^2}\right\}\right] \\
&= \sum_{i=0}^{M}\left[-\frac{1}{2}\ln\left(2\pi s^2\right) - \frac{w_i^2}{2s^2}\right] \\
&= -\frac{M+1}{2}\ln\left(2\pi s^2\right) - \sum_{i=0}^{M}\frac{w_i^2}{2s^2}
\end{aligned} \tag{12}$$

Using equation (12) to equation (10),

$$\begin{aligned}
-\ln p(\mathbf{w}|\mathcal{D}) &= -\ln p(\mathcal{D}|\mathbf{w}) + \frac{M+1}{2}\ln\left(2\pi s^2\right) + \sum_{i=0}^{M}\frac{w_i^2}{2s^2} + \ln p(\mathcal{D}) \\
&= -\ln p(\mathcal{D}|\mathbf{w}) + \sum_{i=0}^{M}\frac{w_i^2}{2s^2} + C \qquad (13)\\
&\qquad\qquad\left[\frac{M+1}{2}\ln\left(2\pi s^2\right) + \ln p(\mathcal{D}) = const. = C\right]
\end{aligned}$$

We know,

$$\ln p(\mathcal{D}|\mathbf{w}) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi) \tag{14}$$

Using equation (14) to (13),

$$-\ln p(\mathbf{w}|\mathcal{D}) = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{N}{2}\ln\sigma^2 + \frac{N}{2}\ln(2\pi) + \sum_{i=0}^{M}\frac{w_i^2}{2s^2} + C$$

Since, $\frac{N}{2}\ln\sigma^2$, $\frac{N}{2}\ln(2\pi)$ and $C$ are not dependent on $\mathbf{w}$, and have no role to play for the maximization of the posterior distribution, we can write,

$$-\ln p(\mathbf{w}|\mathcal{D}) = E(\mathbf{w}) = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{1}{2s^2}\mathbf{w}^T\mathbf{w} \qquad \text{(Ans.)}$$

## II. BAYESIAN ESTIMATION OF A SATELLITE'S ORBITAL FREQUENCY

### 1. Maximum likelihood estimate (MLE)

Given,

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

**(a)** Since $\theta_i \sim \mathcal{N}(\omega t_i, \sigma^2)$, the likelihood function is

$$p\left(\{\theta_i\}_{i=1}^N|\omega\right) = \prod_{i=1}^{N}\mathcal{N}\left(\theta_i|\omega t_i,\sigma^2\right) = \prod_{i=1}^{N}\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{(\theta_i-\omega t_i)^2}{2\sigma^2}\right\} \tag{15}$$

(Ans.)

**(b)** Taking negative natural logarithm on the both sides of equation (15), we get

$$-\ln p\left(\{\theta_i\}_{i=1}^N|\omega\right) = \sum_{i=1}^{N}\left(\frac{1}{2}\ln\left(2\pi\sigma^2\right) + \frac{(\theta_i-\omega t_i)^2}{2\sigma^2}\right) \tag{16}$$

To maximize equation (16) with respect to $\omega$, the first derivative of the equation should be set to zero.

$$\begin{aligned}
&\frac{\partial}{\partial \omega}\left(-\ln p\left(\{\theta_i\}_{i=1}^{N} | \omega\right)\right) = 0 \\
\Rightarrow &\frac{\partial}{\partial \omega}\left(\sum_{i=1}^{N}\left(\frac{1}{2}\ln\left(2\pi\sigma^2\right) + \frac{(\theta_i - \omega t_i)^2}{2\sigma^2}\right)\right) = 0 \\
\Rightarrow &\frac{1}{2\sigma^2}\sum_{i=1}^{N} 2\left(\theta_i - \omega t_i\right)\left(-t_i\right) = 0 \\
\Rightarrow &-\sum_{i=1}^{N} t_i\theta_i + \omega\sum_{i=1}^{N} t_i^2 = 0 \\
\Rightarrow &\omega = \frac{\sum_{i=1}^{N} t_i\theta_i}{\sum_{i=1}^{N} t_i^2} = \hat{\omega}_{MLE} \quad \text{(Showed)}
\end{aligned} \tag{17}$$

**(c)** $\hat{\omega}_{MLE} = 0.01049339792872619\ rad/s$ (Ans.)

**(d)** Noisy $\theta$ and $\hat{\omega}_{MLE}t$ are plotted for comparison.

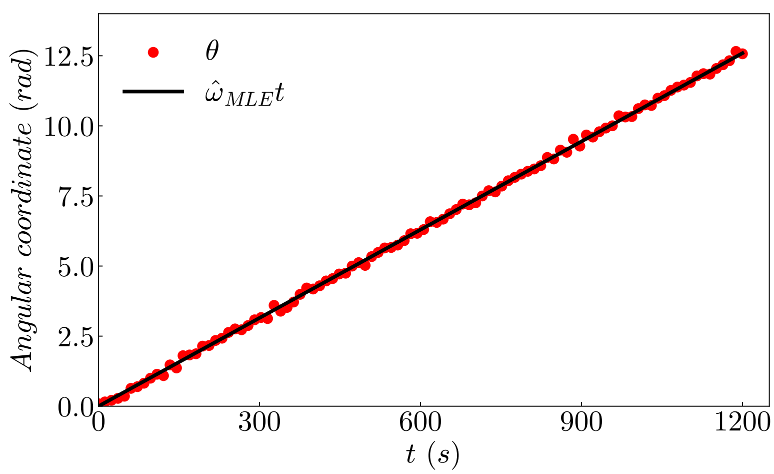



Figure 1: Plot of $\hat{\omega}_{MLE}t$ and $\theta$ vs. $t$.

**(e)** The noise variance $\sigma^2 = 0.007588712373669928\ rad^2$ (Ans.)

**2. Bayesian posterior**

Gaussian prior on $\omega$

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

Gaussian posterior distribution on $\omega$

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

where,

$$\tau_N^2 = \left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\sum_{i=1}^{N} t_i^2\right)^{-1} \tag{18}$$

$$\mu_N = \tau_N^2 \left(\frac{\mu_0}{\tau_0^2} + \frac{1}{\sigma^2}\sum_{i=1}^{N} t_i\theta_i\right) \tag{19}$$

**(a)** Selected prior for mean and variance:

$$\mu_0 = 0, \tau_0^2 = 49$$

$\omega \sim \mathcal{N}(0, 49)$ is chosen for a few reasons:

- Taking the prior mean to be zero ensures no directional bias and symmetry as this is standard Gaussian. This also makes the posterior mean dominated by data only. This prior of mean is just an uninformative reference point.
- Since, wide prior distribution is desired, the prior variance should reflect that. The standard deviation is 7, which makes the distribution flat and spread over a large region so that the posterior is less influenced by the prior. From equation (18), it is easy to see, if $\tau_0^2 >>$, then $\frac{1}{\sigma^2}\sum_{i=1}^{N} t_i^2 >> \frac{1}{\tau_0^2}$. So, the posterior variance is dominated by the data. The same can be understood from equation (19), which becomes $\mu_N = \left(\frac{\tau_N}{\sigma}\right)^2 \sum_{i=1}^{N} t_i\theta_i$ suggesting the dominance of data for posterior mean as well.

The main priority is to not let the priors skew the result away from the actual result underlying beneath the data and let the data dominate in the computation of the posterior.

**(b)** Bayesian posterior mean: $\mu_N = 0.0104933979286925\ rad/s$ (Ans.)

Bayesian posterior variance: $\tau_N^2 = 1.5730371126275355 \times 10^{-10}\ (rad/s)^2$ (Ans.)

**(c)** Bayesian posterior mean, $\mu_N = 0.0104933979286925\ rad/s$

MLE estimate $\hat{\omega}_{MLE} = 0.01049339792872619\ rad/s$

The values are the same till 12 decimal places. So, they are essentially the same. This can be explained through equation (19). If we replace $\mu_0 = 0$ in equation (19), we get

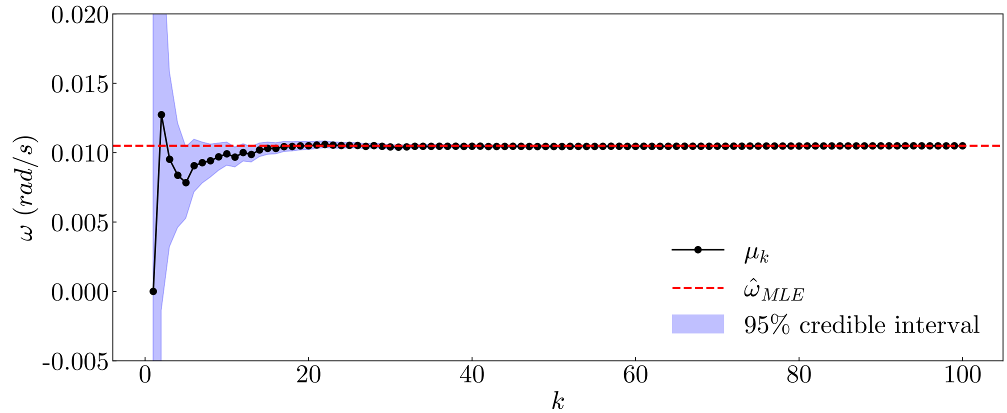
$$\begin{aligned}
\mu_N = \frac{\tau_N^2}{\sigma^2}\sum_{i=1}^{N} t_i\theta_i &= \frac{\left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\sum_{i=1}^{N} t_i^2\right)^{-1}}{\sigma^2}\hat{\omega}_{MLE}\sum_{i=1}^{N} t_i^2 \\
&\qquad\qquad \text{[using equation (17) and (18)]} \\
&\approx \frac{1}{\sigma^2 \cdot \frac{1}{\sigma^2}\sum_{i=1}^{N} t_i^2}\hat{\omega}_{MLE}\sum_{i=1}^{N} t_i^2 \qquad \left[\text{since, } \frac{1}{\sigma^2}\sum_{i=1}^{N} t_i^2 >> \frac{1}{\tau_0^2}\right] \\
&\approx \hat{\omega}_{MLE}
\end{aligned}$$

So, the maximum likelihood solution and Bayesian posterior solution of angular frequency giving the same result further justifies the choice of prior for mean and variance to correctly compute Bayesian posterior.

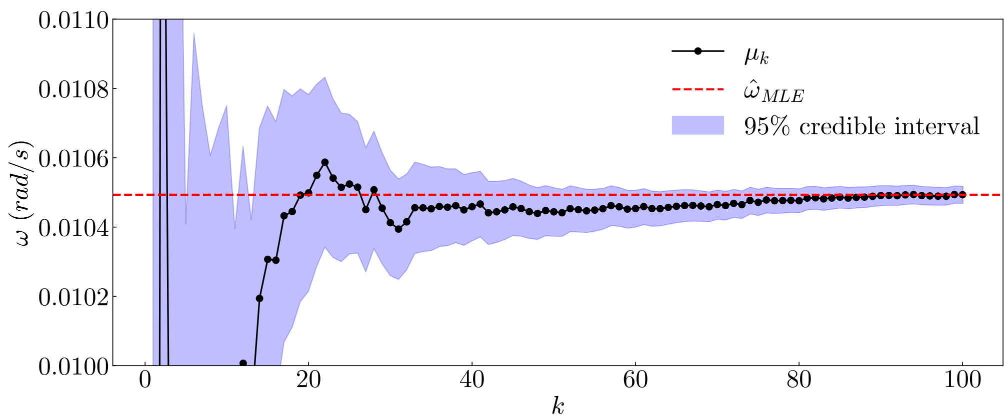
3. **Bayesian filtering**

**(a)** Please refer to listing 1.

**(b)** The uncertainty (band width) shrinks drastically for earlier data points, e.g., $k < 10$. For $k > 10$, the uncertainty keeps reducing, but in a gradual manner, and it continues to do so till $k = 100$. The decay of uncertainty is exponential-like.




(a)




(b)

Figure 2: (a) Sequential posterior mean of angular frequency with respect to no. of observations. (b) Zoomed-in view for better visibility of 95% credible band.

## 4. Information gain over time

Given,

$$H\left(\mathcal{N}\left(\mu,\tau^2\right)\right)=\frac{1}{2}+\frac{1}{2}\ln\left(2\pi\tau^2\right) \tag{20}$$

**(a)** Please refer to listing 1

**(b)** Update of posterior's sequential entropy:

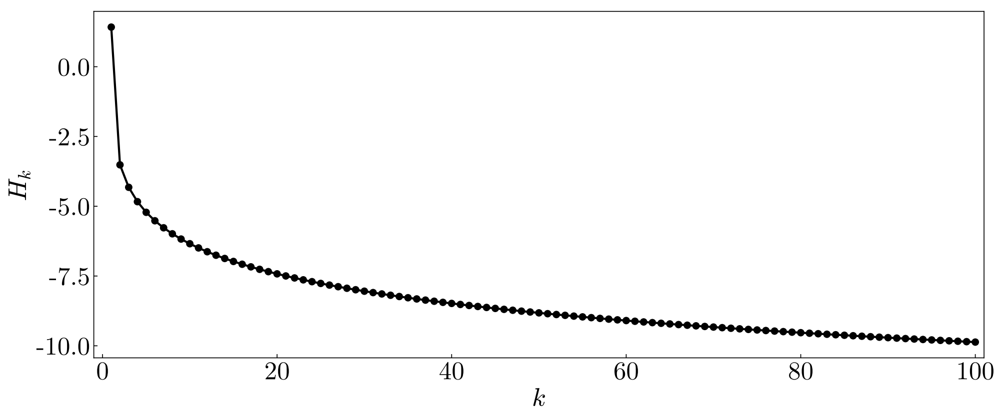



Figure 3: Posterior's entropy with respect to no. of observations.

**(c)** Update of posterior's sequential entropy change:

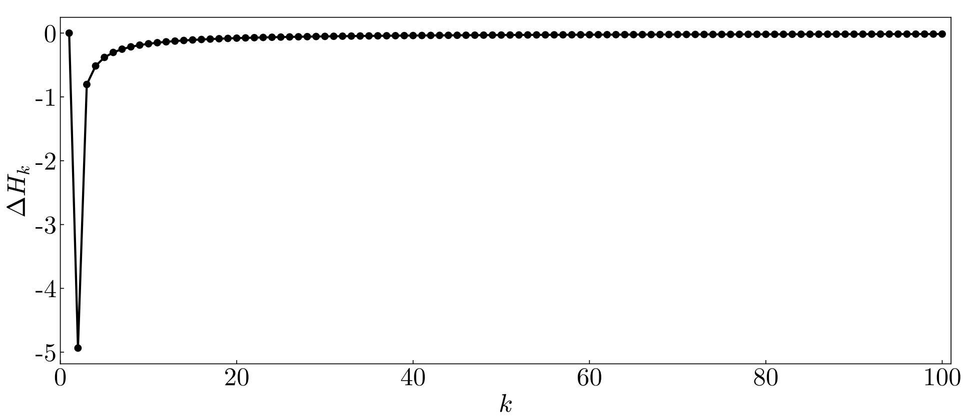



Figure 4: Posterior's entropy change with respect to no. of observations.

**(d)** Entropy only depends on $\tau_k^2$ which can be inferred from equation (20). If we look at how $\tau_k^2$ decreases in figure 5, then we can easily understand how that translates to the change in entropy in figure 3 as well. Initially, there is higher information gain in earlier data with the uncertainty of prior being comparatively large. Once the model sees enough data, enough information accumulates, resulting in very small consecutive changes in entropy.

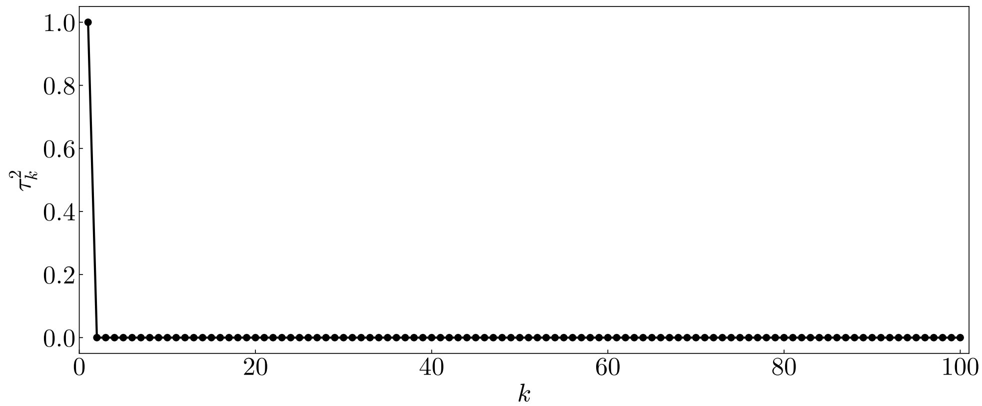



Figure 5: Posterior variance with respect to no. of observations.

From mathematical point of view,

$$\Delta H_k = H_k - H_{k-1} = \frac{1}{2} + \frac{1}{2}\ln\left(2\pi\tau_k^2\right) - \frac{1}{2} - \frac{1}{2}\ln\left(2\pi\tau_{k-1}^2\right) = \frac{1}{2}\ln\left(\frac{\tau_k^2}{\tau_{k-1}^2}\right)$$

Since, $\frac{\tau_k^2}{\tau_{k-1}^2} << 1$ for earlier data points, $\Delta H_k$ also reflects that by getting reduced rapidly. So, it can be concluded that with higher uncertainty in prior, the posterior precision increases steeply because of higher information gain. Once information accumulates, the information gain over new observations diminishes.

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl

# read csv
data = np.loadtxt("hw3.csv", delimiter=",")
t = data[:,0]
theta = data[:,1]

# ===== Problem-1 =====

# compute maximum likelihood solution
omega_mle = np.sum(t*theta)/np.sum(t*t)
#print(np.sum(t*theta))
#print(np.sum(t*t))
print("1(c) Maximum likelihood solution: ", omega_mle)

# parameters for plotting
plt.rcParams['font.family'] = 'serif'
plt.rcParams['font.serif'] = ['CMU Serif']
plt.rcParams['mathtext.fontset'] = 'cm'
plt.rcParams['font.size'] = 24
mpl.rcParams['axes.unicode_minus'] = False

# plot
fig, ax = plt.subplots(figsize=(10, 6))

ax.scatter(t, theta, color='red', label=r"$\theta$", linewidth = 3)
```

```python
ax.plot(t, omega_mle*t, linestyle='-', color='black', label=r"$\hat\omega_{MLE} t$", linewidth = 3)

plt.xlim(0, 1250)
ax.set_xticks(np.arange(0, 1250, 300))
plt.ylim(0, 14)
plt.xlabel(r"$t$ ($s$)")
plt.ylabel("$Angular$ $coordinate$ ($rad$)")
plt.legend(loc='upper left', frameon=False)
plt.tick_params(axis="both", which="both", direction="in")
plt.savefig(f'1d.pdf', dpi=1080)
plt.show()

# compute noise variance
var = 1/(len(t)-1)*np.sum((theta-omega_mle*t)**2)
print("1(e) Noise variance: ", var)

# ===== Problem-2 =====

# selected priors for mean and variance
mu_0 = 0
tau_0_sq = 49

# compute posterior variance
tau_n_sq = 1/((1/tau_0_sq)+((1/var)*np.sum(t**2)))
print("2(b) Bayesian posterior variance: ", tau_n_sq)

# compute posterior mean
mu_n = tau_n_sq*((mu_0/tau_0_sq) + (1/var)*np.sum(t*theta))
print("2(b) Bayesian posterior mean: ", mu_n)

# ===== Problem-3 =====

# initialize posterior mean, variance, $\sum t^2$, $sum t\theta$
mu = np.zeros(len(t))
tau_sq = np.zeros(len(t))
s_tt = 0.0
s_ttheta = 0.0

# sequential update of posterior mean and variance
for k in range(len(t)):
    s_tt += t[k]**2
    s_ttheta += t[k]*theta[k]
    tau_sq_ = 1/((1/tau_0_sq)+((1/var)*s_tt))
    mu_ = tau_sq_*((mu_0/tau_0_sq) + (1/var)*s_ttheta)
    tau_sq[k] = tau_sq_
    mu[k] = mu_

print("3(a) mu_k: ", mu)
print("3(a) tau_sq_k: ", tau_sq)

# plot
fig, ax = plt.subplots(figsize=(15, 6))
plt.plot(np.arange(1, len(t)+1), mu, "-o", color="black", label=r'$\mu_k$')
plt.axhline(omega_mle, color="red", linestyle='--', linewidth=2, label=r'$\hat{\omega}_{MLE}$')
plt.fill_between(np.arange(1, len(t)+1), mu-1.96*np.sqrt(tau_sq), mu+1.96*np.sqrt(tau_sq),
                 color="blue", alpha=0.25, rasterized=True, label="95% credible
```

```
                    interval")
plt.xlabel(r"$k$")
plt.ylabel(r"$\omega$ $(rad/s)$")
plt.legend(loc="lower right", frameon=False)
plt.tick_params(axis="both", which="both", direction="in")
plt.ylim(-0.005, 0.02)
plt.savefig(f'3b-1.pdf', dpi=1080)
plt.show()

# zoomed plot
fig, ax = plt.subplots(figsize=(15, 6))
plt.plot(np.arange(1, len(t)+1), mu, "-o", color="black", label=r'$\mu_k$')
plt.axhline(omega_mle, color="red", ls='--', linewidth=2, label=r'$\hat{\omega}_{MLE}$')
plt.fill_between(np.arange(1, len(t)+1), mu-1.96*np.sqrt(tau_sq), mu+1.96*np.sqrt(tau_sq),
                 color="blue", alpha=0.25, rasterized=True, label="95% credible interval")
plt.xlabel(r"$k$")
plt.ylabel(r"$\omega$ $(rad/s)$")
plt.legend(loc="upper right", frameon=False)
plt.tick_params(axis="both", which="both", direction="in")
plt.ylim(0.01, 0.011)
plt.savefig(f'3b-2.pdf', dpi=1080)
plt.show()

# ===== Problem-4 =====

# set prior for variance an entropy
tau_0_sq_ = 1
h_0 = 1/2+1/2*np.log(2*np.pi*tau_0_sq_)

# reinitialize variance, $\sum t^2$; initialize entropy and difference in entropy
tau_sq = np.zeros(len(t))
h = np.zeros(len(t))
dh = np.zeros(len(t))
s_tt = 0.0

# sequential update of entropy and difference in entropy
for k in range(len(t)):
    s_tt += t[k]**2
    tau_sq_ = 1/((1/tau_0_sq_)+((1/var)*s_tt))
    h_ = 1/2+1/2*np.log(2*np.pi*tau_sq_)
    if k == 0:
        dh_ = h_ - h_0
    else:
        dh_ = h_ - h[k-1]
    tau_sq[k] = tau_sq_
    h[k] = h_
    dh[k] = dh_

print("4(a) tau_sq_k: ", tau_sq)
print("4(a) H_k: ", h)
print("4(c) dH_k: ", dh)

# plot
fig, ax = plt.subplots(figsize=(15, 6))

plt.plot(np.arange(1, len(t)+1), h, "-o", linewidth=2, color="black")
```

```python

plt.xlabel(r"$k$")
plt.ylabel(r"$H_k$")
plt.xlim(-1, 101)
plt.tick_params(axis="both", which="both", direction="in")
plt.savefig(f'4b.pdf', dpi=1080)
plt.show()

fig, ax = plt.subplots(figsize=(15, 6))

plt.plot(np.arange(1, len(t)+1), dh, "-o", linewidth=2, color="black")

plt.xlabel(r"$k$")
plt.ylabel(r"$\Delta H_k$")
plt.xlim(0, 101)
plt.tick_params(axis="both", which="both", direction="in")
plt.savefig(f'4c.pdf', dpi=1080)
plt.show()

fig, ax = plt.subplots(figsize=(15, 6))

plt.plot(np.arange(1, len(t)+1), tau_sq, "-o", linewidth=2, color="black")

plt.xlabel(r"$k$")
plt.ylabel(r"$\tau_k^2$")
plt.xlim(0, 101)
plt.tick_params(axis="both", which="both", direction="in")
plt.savefig(f'4d.pdf', dpi=1080)
plt.show()
```

Listing 1: *bayes_frequency.py*

## III. BAYESIAN ESTIMATION OF A NEWLY DISCOVERED STAR'S MASS

Estimated mass of star using Monte Carlo sampling, $M = 2.3091475263003872 \times 10^{30}\ kg$

Since $M \sim \omega^2$, the distribution of $M$ should be skewed because of nonlinearity, but since the posterior ratio of $\mu_N$ and $\tau_N$ of $\omega$ is 21.25, the distribution of $M$ turns out to be almost exactly Gaussian from figure 7.

Error bar, $M \pm \delta M = (2.3091475263003872 \pm 0.42518843856163555) \times 10^{30}\ kg$

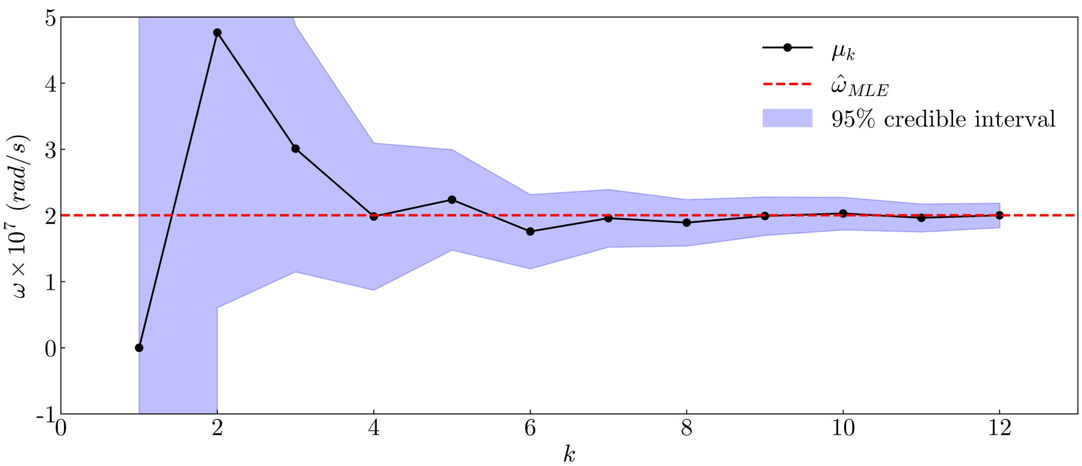



Figure 6: Posterior mean of angular frequency with respect to no. of observations.

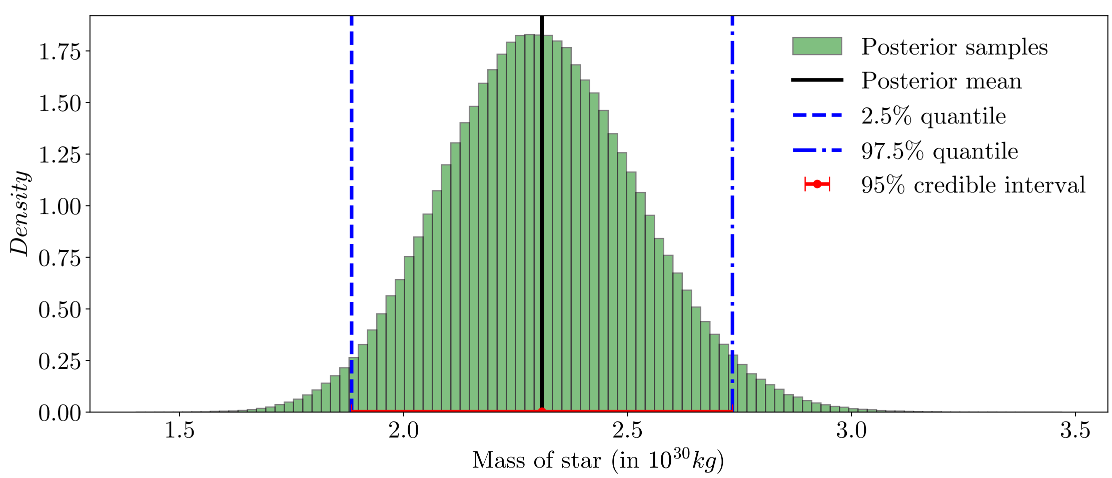



Figure 7: Distribution of the star's mass derived from posterior angular frequency using Monte Carlo sampling.

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl

# read csv
data = np.loadtxt("SolarData-Hw3.csv", delimiter=",")
t = data[:,0]
print(t)
theta = data[:,1]

# prior for mean and variance
mu_0 = 0.0
tau_0_sq = 49

# maximum likelihood and noise variance
omega_mle = np.sum(t*theta)/np.sum(t*t)
print("MLE = ", omega_mle)
var = 1/(len(t)-1)*np.sum((theta-omega_mle*t)**2)

# constants
G = 6.6743e-11
r = 1.56479e11

# initialize posterior mean, variance, $\sum t^2$, $sum t\theta$
mu = np.zeros(len(t))
tau_sq = np.zeros(len(t))
s_tt = 0.0
s_ttheta = 0.0

# sequential update of posterior mean and variance
for k in range(len(t)):
    s_tt += t[k]**2
    s_ttheta += t[k]*theta[k]
    tau_sq_ = 1/((1/tau_0_sq)+((1/var)*s_tt))
    mu_ = tau_sq_*((mu_0/tau_0_sq) + (1/var)*s_ttheta)
    tau_sq[k] = tau_sq_
    mu[k] = mu_

print("Bayesian posterior mean = ", mu[-1])
print("Bayesian posterior variance = ", tau_sq[-1])
print("Bayesian posterior mean/std. deviation = ", mu[-1]/np.sqrt(tau_sq[-1]))

# parameters for plotting
plt.rcParams['font.family'] = 'serif'
plt.rcParams['font.serif'] = ['CMU Serif']
plt.rcParams['mathtext.fontset'] = 'cm'
plt.rcParams['font.size'] = 20
mpl.rcParams['axes.unicode_minus'] = False

# plot
fig, ax = plt.subplots(figsize=(15, 6))
ax.plot(np.arange(1, len(t)+1), mu*1e7, "-o", color="black", label=r'$\mu_k$')
ax.axhline(omega_mle*1e7, color="red", linestyle='--', linewidth=2, label=r'$\hat{\omega}_{MLE}$')
ax.fill_between(np.arange(1, len(t)+1), (mu-1.96*np.sqrt(tau_sq))*1e7, (mu+1.96*np.sqrt(tau_sq))*1e7,
                 color="blue", alpha=0.25, rasterized=True, label="95% credible interval")
```

```python
plt.xlabel(r"$k$")
plt.ylabel(r"$\omega \times 10^7$ $(rad/s)$")
plt.legend(loc="upper right", frameon=False)
plt.tick_params(axis="both", which="both", direction="in")
plt.xlim(0, 13)
plt.ylim(-1, 5)
plt.savefig(f'omega_bayes.pdf', dpi=1080)
plt.show()

# Monte Carlo sampling
np.random.seed(32) # for reproducibility
n = 1000000 # no. of realizations
omega_smpl = np.random.normal(mu[-1], np.sqrt(tau_sq[-1]), n)
mass_smpl = (r**3 / G) * (omega_smpl**2)

# mean and bounds
mass_mean = np.mean(mass_smpl)
mass_var = np.var(mass_smpl, ddof=1)
#mass_low, mass_high = np.percentile(mass_smpl, [2.5, 97.5])

print(f"Posterior mean mass: {mass_mean}")
print(f"95% credible interval: [{mass_mean} - {1.96*np.sqrt(mass_var)}, {mass_mean} + {1.96*np.sqrt(mass_var)}]")
#print(f"95% credible interval: [{mass_mean} - {mass_mean - mass_low}, {mass_mean} + {-mass_mean + mass_high}] kg")

# plot of sampless mass
fig, ax = plt.subplots(figsize=(15, 6))

ax.hist(mass_smpl/1e30, bins=100, color='green', edgecolor='black', alpha=0.5, density=True, label='Posterior samples')

# 95% credible interval
ax.axvline(mass_mean/1e30, color='black', lw=3, label='Posterior mean')
ax.axvline((mass_mean-1.96*np.sqrt(mass_var))/1e30, color='blue', lw=3, linestyle='--', label='2.5% quantile')
ax.axvline((mass_mean+1.96*np.sqrt(mass_var))/1e30, color='blue', lw=3, linestyle='-.', label='97.5% quantile')

# errorbar
ax.errorbar(mass_mean/1e30, 0.002, xerr=[[1.96*np.sqrt(mass_var)/1e30],
            [1.96*np.sqrt(mass_var)/1e30]], fmt='o', color='red', capsize=6,
            lw=3, label='95% credible interval')

ax.set_xlabel(r"Mass of star (in $10^{30}kg$)")
ax.set_ylabel(r"$Density$")
ax.legend(frameon=False)
plt.savefig(f'hist.pdf', dpi=1080)
plt.show()
```

Listing 2: *solar.py*