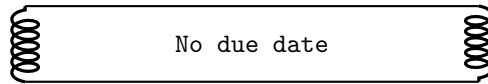


Exercise 3



Instruction: These are for your own practice. There is no need to submit your answer.

Question 1

Observations $X_1, \dots, X_n$ are independent identically distributed, following the PDF $f_{X_i}(x_i) = \theta x_i^{\theta-1}$, and that $0 < X_i < 1$ for all i . The parameter θ is an unknown positive number. Find the ML estimator of θ .

The log-likelihood function is given by

$$\log f_X(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \log \theta x_i^{\theta-1} = n \log \theta + (\theta-1) \sum_{i=1}^n \log x_i$$

Now we can take the derivative (w.r.t. θ) and set to zero, such that

$$\begin{aligned} \frac{n}{\theta} + \sum_{i=1}^n \log x_i &= 0 \\ \theta &= -\frac{n}{\log(x_1 \dots x_n)} \end{aligned}$$

Hence, the ML estimate is

$$\hat{\Theta}_n = -\frac{n}{\log(X_1 \dots X_n)} = -\frac{n}{\sum_{i=1}^n \log X_i}$$

Question 2

Observations $X_1, \dots, X_n$ are independent identically distributed, being Rayleigh with $f_X(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$, where σ^2 is unknown. Find the ML estimator of σ^2 .

The likelihood function is

$$\begin{aligned} f_X(x_1, \dots, x_n; \sigma^2) &= f_{X_1}(x_1; \sigma^2) \cdots f_{X_n}(x_n; \sigma^2) \\ &= \prod_{i=1}^n \frac{x_i}{\sigma^2} e^{-x_i^2/(2\sigma^2)} \end{aligned}$$

The log-likelihood function is

$$\begin{aligned} \log f_X(x_1, \dots, x_n; \sigma^2) &= \log \prod_{i=1}^n \frac{x_i}{\sigma^2} e^{-x_i^2/(2\sigma^2)} \\ &= \sum_{i=1}^n \left(\log x_i - \log \sigma^2 - \frac{x_i^2}{2\sigma^2} \right) \end{aligned}$$

Let $s = \sigma^2$. Now, we take derivative with respect to s , and set to zero:

$$\begin{aligned} \frac{d \log f_X(x_1, \dots, x_n; \sigma^2)}{ds} &= \frac{d}{ds} \sum_{i=1}^n \left(\log x_i - \log s - \frac{x_i^2}{2} s^{-1} \right) \\ &= \sum_{i=1}^n \left(-s^{-1} + \frac{x_i^2}{2} s^{-2} \right) \\ &= \frac{1}{s} \sum_{i=1}^n \left(-1 + \frac{x_i^2}{2} s^{-1} \right) = 0 \end{aligned}$$

This gives the ML estimate

$$\begin{aligned} -n + \frac{\sum_{i=1}^n x_i^2}{2} \frac{1}{s} &= 0 \\ \hat{\sigma}^2 = \hat{s} &= \frac{1}{2n} \sum_{i=1}^n x_i^2 \end{aligned}$$

Question 3

In a certain production line, it is known that the lifetime (measured in hours) of the products follows a Poisson distribution with parameter λ . Five such products were sampled and tested, and they lasted, respectively, 5.2, 5.3, 5.5, 5.7, 5.3 hours. Find the ML estimator of λ .

As derived in the course notes, the ML estimator of a Poisson distribution is the average of its observed values. Hence,

$$\hat{\lambda} = \frac{5.2 + 5.3 + 5.5 + 5.7 + 5.3}{5} = 5.4$$

Question 4

Observations $X_1, \dots, X_n$ are independent identically distributed, following the Gamma PDF given by

$$f_{X_i}(x_i) = \frac{\beta^\alpha x_i^{\alpha-1} e^{-\beta x_i}}{\Gamma(\alpha)},$$

where α and β are parameters of the distribution, and $\Gamma(\cdot)$ is the Gamma function (see Lecture 1 if you don't remember what this function is), with derivative denoted by $\Gamma'(\cdot)$. We wish to obtain the ML estimators of α and β , respectively denoted by $\hat{\alpha}$ and $\hat{\beta}$, based on the observations.

1. Show that, by maximizing the log-likelihood, we have the relationship

$$\hat{\beta} = \frac{\hat{\alpha}}{\frac{1}{n} \sum x_i}$$

where the summation ranges from $i = 1$ to $i = n$.

2. Show that, by maximizing the log-likelihood and using the relationship above, we have

$$\log \hat{\alpha} - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = \log \left(\frac{1}{n} \sum x_i \right) - \frac{1}{n} \sum \log x_i$$

The likelihood function is given by

$$f_X(x_1, \dots, x_n; \alpha, \beta) = \frac{\beta^{n\alpha} (\prod x_i^{\alpha-1}) e^{-\beta \sum x_i}}{[\Gamma(\alpha)]^n}$$

The log-likelihood function is given by

$$\log f_X(x_1, \dots, x_n; \alpha, \beta) = n\alpha \log \beta + (\alpha - 1) \sum \log x_i - \beta \sum x_i - n \log \Gamma(\alpha)$$

1. Now we can first take the derivative (w.r.t. β) and set to zero, such that

$$\frac{n\alpha}{\beta} - \sum x_i = 0$$

Hence, the ML estimate is

$$\hat{\beta} = \frac{\hat{\alpha}}{\frac{1}{n} \sum x_i}$$

2. Now we can take the derivative (w.r.t. α) and set to zero, such that

$$n \log \beta + \sum \log x_i - \frac{n}{\Gamma(\alpha)} \Gamma'(\alpha) = 0$$

Hence, the ML estimate is

$$\begin{aligned} \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} &= \log \hat{\beta} + \frac{1}{n} \sum \log x_i \\ &= \log \hat{\alpha} - \log \left(\frac{1}{n} \sum x_i \right) + \frac{1}{n} \sum \log x_i \\ \log \hat{\alpha} - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} &= \log \left(\frac{1}{n} \sum x_i \right) - \frac{1}{n} \sum \log x_i \end{aligned}$$

Question 5

Observations $X_1, X_2, \dots, X_n$, where $n > 2$, are independent identically distributed and drawn from a normal distribution $N \sim (0, \sigma^2)$ with $\sigma > 0$.

1. Show that the estimator $\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ is an unbiased estimator of σ^2 .
2. Let $Y \sim N(\mu, \sigma^2)$. Show that $\mathbb{E}((Y - \mu)^4) = 3\sigma^4$.
Hint: Let $Z = (Y - \mu)/\sigma$. What is $\mathbb{E}(Z^4)$ and how can this help with solving this part? You may need to evaluate some integrals directly.
3. Compute $\text{var}((Y - \mu)^2)$.

1. We have

$$\mathbb{E}(\tilde{\sigma}_n^2) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2) = \sigma^2$$

So this is an unbiased estimator.

2. Let $Z = \frac{Y - \mu}{\sigma}$, and it follows that $Z \sim N(0, 1)$. Thus,

$$\begin{aligned} \mathbb{E}\left(\left(\frac{Y - \mu}{\sigma}\right)^4\right) &= \mathbb{E}(Z^4) \\ &= \int_{-\infty}^{\infty} z^4 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= -z^3 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Big|_{-\infty}^{\infty} + 3 \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= 3\mathbb{E}(Z^2) \\ &= 3\text{var}(Z) \\ &= 3 \end{aligned}$$

It follows that $\mathbb{E}((Y - \mu)^4) = 3\sigma^4$.

3. We have

$$\begin{aligned} \text{var}((Y - \mu)^2) &= \mathbb{E}((Y - \mu)^4) - \mathbb{E}((Y - \mu)^2)^2 \\ &= 3\sigma^4 - (\sigma^2)^2 \\ &= 2\sigma^4. \end{aligned}$$

Question 6

A researcher is conducting an experiment. The probability that he will succeed is a continuous random variable denoted by X , where

$$f_X(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

During a research project, he will continue this experiment for Y times until he succeeds for the first time. Now we know that he has succeeded in the third attempt. Find the MAP estimate of X .

We know that $Y|X$ follows a geometric random variable $\text{Geo}(x)$, and therefore

$$f_{Y|X}(y|x) = x(1-x)^{y-1}$$

for $y = 1, 2, \dots$

Using Bayes' rule, we have

$$\begin{aligned} f_{X|Y}(x|y) &\propto f_{Y|X}(y|x) f_X(x) \\ &= x(1-x)^{3-1} \cdot 2x \\ &= 2x^2(1-x)^2 \end{aligned}$$

where $0 \leq x \leq 1$. Differentiating with respect to x and setting the result to zero, we have

$$\frac{d}{dx}[x^2(1-x)^2] = 4x(1-x)^2 - 4(1-x)x^2 = 0.$$

Solving for x , we obtain the MAP estimate as

$$\hat{x} = \frac{1}{2}.$$

Question 7

Observations $X = \{X_1, \dots, X_n\}$ are independent identically distributed, being Gaussian with $\mathcal{N}(\mu, \sigma^2)$, where μ is unknown while σ^2 is known. Suppose that μ itself (i.e. the prior distribution) is a Gaussian random variable with mean ν and variance β^2 . We would like to find and analyze the MAP estimate of μ .

1. Let $C = f_X(x)$. Using Bayes' rule, show that the posterior probability can be written as

$$f_{\mu|X}(\mu|x) = \frac{1}{C} \left(\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \left(\frac{1}{\sqrt{2\pi}\beta} e^{-\frac{(\mu - \nu)^2}{2\beta^2}} \right)$$

2. Show that

$$\sum_{i=1}^N \frac{x_i - \mu}{\sigma^2} = \frac{\mu - \nu}{\beta^2}$$

by maximizing the log posterior.

3. By rearranging the terms, show that

$$\hat{\mu} = \frac{\beta^2 \sum_{i=1}^N x_i + \sigma^2 \nu}{\beta^2 N + \sigma^2}$$

4. What happens when N grows large? Note that we have

$$\mu_{ML} = \frac{1}{N} \sum_{i=1}^N x_i$$

which is the maximum likelihood estimate of μ .

1. Bayes' rule allows us to write

$$f_{\mu|X}(\mu|x) = \frac{f_{X|\mu}(x|\mu) f_{\mu}(\mu)}{f_X(x)} = \frac{1}{C} \left(\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \left(\frac{1}{\sqrt{2\pi}\beta} e^{-\frac{(\mu - \nu)^2}{2\beta^2}} \right)$$

2. The log posterior is

$$\log f_{\mu|X}(\mu|x) = \sum_{i=1}^N \left(-\log \sqrt{2\pi}\sigma - \frac{(x_i - \mu)^2}{2\sigma^2} \right) - \log \sqrt{2\pi}\beta - \frac{(\mu - \nu)^2}{2\beta^2} - \log C$$

Taking the derivative with respect to μ and setting the result to zero, we have

$$\frac{\partial \log f_{\mu|X}(\mu|x)}{\partial \mu} = \sum_{i=1}^N \frac{x_i - \mu}{\sigma^2} - \frac{\mu - \nu}{\beta^2} = 0$$

Therefore,

$$\sum_{i=1}^N \frac{x_i - \mu}{\sigma^2} = \frac{\mu - \nu}{\beta^2}$$

3. We group together the terms involving μ :

$$\begin{aligned} \frac{\sum_{i=1}^N x_i}{\sigma^2} - \frac{N}{\sigma^2} \mu &= \frac{1}{\beta^2} \mu - \frac{\nu}{\beta^2} \\ \frac{\sum_{i=1}^N x_i}{\sigma^2} + \frac{\nu}{\beta^2} &= \left(\frac{N}{\sigma^2} + \frac{1}{\beta^2} \right) \mu \\ \hat{\mu} &= \frac{\beta^2 \sum_{i=1}^N x_i + \sigma^2 \nu}{\beta^2 N + \sigma^2} \end{aligned}$$

4. We can write the MAP estimate as

$$\hat{\mu} = \frac{\frac{1}{N} \sum_{i=1}^N x_i + \frac{\sigma^2 \nu}{\beta^2 N}}{1 + \frac{\sigma^2}{\beta^2 N}}$$

As N grows larger, the second terms of both the numerator and the denominator approaches zero. Therefore,

$$\lim_{N \rightarrow \infty} \hat{\mu} = \mu_{ML}$$

As we have more and more data, the prior information becomes increasingly unimportant!

Question 8

The photon arrival at a particular photodetector is modeled as a Poisson distribution, with $X \sim \text{Poisson}(\theta)$, i.e.,

$$f_X(x) = \frac{e^{-\theta} \theta^x}{x!},$$

where θ is an unknown to be estimated from some observation data. We assume that θ follows a Gamma distribution denoted as $\Theta \sim \text{Gamma}(\alpha, \beta)$, given by

$$f_{\Theta}(\theta; \alpha, \beta) = \frac{\beta^{\alpha} \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)},$$

where α and β are parameters of the distribution, and $\Gamma(\cdot)$ is the Gamma function. It is known that the mode of the Gamma distribution (with $\alpha \geq 1$) is given by $(\alpha - 1)/\beta$.

Assume we make a series of observations, X_i for $1 \leq i \leq n$, which are independent identically distributed, following the same Poisson distribution.

1. Show that the posterior is given by

$$f_{\theta|X_1, \dots, X_n}(\theta | x_1, \dots, x_n) = c \cdot e^{-(\beta+n)\theta} \cdot \theta^{(\alpha+s)-1},$$

where $s = \sum_{i=1}^n x_i$, and c is a certain constant.

2. Express the posterior in terms of a Gamma distribution.
3. Find the MAP estimate of θ .
4. What happens to the MAP estimate when n is large?

1. The likelihood function is

$$\begin{aligned} f_X(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} \\ &\propto e^{-n\theta} \theta^{\sum_{i=1}^n x_i} = e^{-n\theta} \theta^s. \end{aligned}$$

The posterior is the product of the likelihood and the prior, i.e.,

$$\begin{aligned} f_{\theta|X_1, \dots, X_n}(\theta | x_1, \dots, x_n) &\propto \frac{\beta^{\alpha} \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)} \cdot e^{-n\theta} \theta^s \\ &\propto e^{-(\beta+n)\theta} \theta^{(\alpha+s)-1} \end{aligned}$$

Therefore,

$$f_{\theta|X_1, \dots, X_n}(\theta | x_1, \dots, x_n) = c \cdot e^{-(\beta+n)\theta} \cdot \theta^{(\alpha+s)-1},$$

where c is a certain constant.

2. Comparing the form of the posterior with the Gamma distribution given above, we can see that

$$\theta | X_1, \dots, X_n \sim \text{Gamma}(\alpha + s, \beta + n).$$

The normalization will work out correctly because this is a PDF.

3. Since the peak (mode) of $\text{Gamma}(\alpha, \beta)$ is $(\alpha - 1)/\beta$, the MAP estimate is therefore

$$\hat{\theta} = \frac{\alpha + s - 1}{\beta + n}$$

4. We can write

$$\hat{\theta} = \frac{\alpha/n + s/n - 1/n}{\beta/n + 1}.$$

When n is large, $\alpha/n \approx 0$, and $\beta/n \approx 0$, and also $1/n \approx 0$, so

$$\hat{\theta} \approx \frac{s}{n}.$$

This is the ML estimate!

Question 9

We have developed an optical imaging system for single-cell inspection. Each cell can be either normal or cancerous. Through extensive studies, we find that the two types of cells have different physiological characteristics. In particular, it is possible to differentiate them in terms of the cell diameter.

For a normal cell, the mean diameter is $\mu_1 = 3 \mu\text{m}$, with standard deviation $\sigma_1 = 0.4 \mu\text{m}$.

For a cancerous cell, the mean diameter is $\mu_2 = 4.5 \mu\text{m}$, with standard deviation $\sigma_2 = 0.6 \mu\text{m}$.

In an experiment, we have a mixture containing 80% of normal and 20% of cancerous cells. We now image a specific cell and find that it has a diameter of $3.7 \mu\text{m}$.

1. We assume that both normal and cancerous cells follow Gaussian distributions, with mean and standard deviation as stated above.
 - (a) What is the posterior probability that the cell is normal?
 - (b) What is the posterior probability that the cell is cancerous?
 - (c) Using MAP, would you conclude the cell is normal or cancerous?
2. On the basis of some new theory, we change our models, and now both cell distributions are Laplacian centered around the mean value stated above. Again, their mean and standard deviation remain unchanged.

- (a) The PDF of normal cell is

$$f_X(x) = \frac{\lambda_1}{2} e^{-\lambda_1 |x - m_1|}.$$

Compute λ_1 and m_1 .

- (b) The PDF of cancerous cell is

$$f_X(x) = \frac{\lambda_2}{2} e^{-\lambda_2 |x - m_2|}.$$

Compute λ_2 and m_2 .

- (c) What is the posterior probability that the cell is normal?
- (d) What is the posterior probability that the cell is cancerous?
- (e) Using MAP, would you conclude the cell is normal or cancerous?

Let J denote the cell type, where $J = 1$ indicates a normal cell, and $J = 2$ indicates a cancerous cell. Furthermore, let L be the cell diameter. From Bayes' rule, we know that

$$f_{J|L}(j|x) = \frac{f_{L|J}(x|j) \mathbb{P}(J = j)}{f_{L|J}(x|1) \mathbb{P}(J = 1) + f_{L|J}(x|2) \mathbb{P}(J = 2)}$$

1. We plug in Gaussian models for the two cell types.

(a) The posterior probability that the cell is normal is given by

$$f_{J|L}(j = 1 | x = 3.7) = \frac{\frac{0.8}{0.4} e^{-(3.7-3)^2/(2 \cdot 0.4^2)}}{\frac{0.8}{0.4} e^{-(3.7-3)^2/(2 \cdot 0.4^2)} + \frac{0.2}{0.6} e^{-(3.7-4.5)^2/(2 \cdot 0.6^2)}} \approx 0.76$$

(b) The posterior probability that the cell is cancerous is given by

$$1 - 0.76 = 0.24$$

(c) Using MAP, we conclude that the cell is normal.

2. Now, we plug in Laplacian models for the two cell types.

(a) $m_1 = 3$, since this is the mean. The variance of a Laplacian random variable is $2/\lambda_1^2$, so $\sqrt{2}/\lambda_1 = 0.4$, or $\lambda_1 = 3.54$.

(b) Similarly, $m_2 = 4.5$, and $\lambda_2 = \sqrt{2}/0.6 = 2.36$.

(c) The posterior probability that the cell is normal is given by

$$f_{J|L}(j = 1 | x = 3.7) = \frac{\frac{0.8}{0.4} e^{-\sqrt{2}|3.7-3|/(0.4)}}{\frac{0.8}{0.4} e^{-\sqrt{2}|3.7-3|/(0.4)} + \frac{0.2}{0.6} e^{-\sqrt{2}|3.7-4.5|/(0.6)}} \approx 0.77$$

(d) The posterior probability that the cell is cancerous is given by

$$1 - 0.77 = 0.23$$

(e) Using MAP, we again conclude that the cell is normal.

Question 10

Imagine that you are a data scientist working for an e-commerce platform. The company has recently implemented a new recommendation system, and you are tasked with evaluating its performance. Specifically, you are interested in the probability, $0 \leq \theta \leq 1$, that a user will click on a recommended product. This probability is unknown and you wish to estimate it using data collected from user interactions.

After monitoring the new recommendation system, you gather a small sample of user interactions: $x = (0, 1, 0, 1, 0)$, where a "1" indicates a click and a "0" indicates no click.

1. You want to estimate the probability of a user clicking on a recommendation using the MLE method. What is the maximum likelihood (ML) estimate for θ based on this data?
2. The marketing team believes that the click probability could only be one of three values based on previous marketing studies: 0.3, 0.5, 0.6. With this new restriction, how does the ML for θ change?
3. To refine the estimate, the team decides to incorporate prior marketing knowledge into the estimation process. They assign different levels of belief (prior probabilities) to each of the three possible values of θ : $\mathbb{P}(\theta = 0.3) = 0.05$, $\mathbb{P}(\theta = 0.5) = 0.01$, and $\mathbb{P}(\theta = 0.6) = 0.94$. How do you compute the MAP for θ under these new assumptions?
4. You then decide to conduct an experiment: you want to demonstrate to the team that by altering the prior probabilities, you can control the outcome of the MAP estimation. How would you set the priors for θ if you want the MAP estimate to be 0.3?
5. As you collect more and more samples, what would be the relationship between using ML and MAP for the estimation? why?

1. The likelihood of drawing samples $\mathbf{x} = (0, 1, 0, 1, 0)$ can be modeled as a Bernoulli distribution,

$$\mathbb{P}(\mathbf{x}; \theta) = \theta^2(1 - \theta)^3$$

The log-likelihood is

$$\log \mathbb{P}(\mathbf{x}; \theta) = 2 \log \theta + 3 \log(1 - \theta)$$

The maximizer can be found by setting its first derivative to zero,

$$\frac{\partial}{\partial \theta} \log \mathbb{P}(\mathbf{x}; \theta) = \frac{2}{\theta} - \frac{3}{1 - \theta} = 0 \Rightarrow \hat{\theta} = \frac{2}{5}$$

2. We need to find out which of the three acceptable θ values maximizes the likelihood, and since there are only finitely many, we can just plug them all in and compare.

$$\mathbb{P}(\mathbf{x}; \theta = 0.3) = 0.3^2 \times 0.7^3 = 0.03087$$

$$\mathbb{P}(\mathbf{x}; \theta = 0.5) = 0.5^2 \times 0.5^3 = 0.03125$$

$$\mathbb{P}(\mathbf{x}; \theta = 0.6) = 0.6^2 \times 0.4^3 = 0.02304$$

So $\hat{\theta} = 0.5$.

3. Instead of maximizing just the likelihood, we need to maximize the likelihood times the prior.

$$\mathbb{P}(\theta = 0.3|\mathbf{x}) \propto \mathbb{P}(\mathbf{x}|\theta = 0.3) \mathbb{P}(\theta = 0.3) = 0.03087 \times 0.05 = 0.001544$$

$$\mathbb{P}(\theta = 0.5|\mathbf{x}) \propto \mathbb{P}(\mathbf{x}|\theta = 0.5) \mathbb{P}(\theta = 0.5) = 0.03125 \times 0.01 = 0.000313$$

$$\mathbb{P}(\theta = 0.6|\mathbf{x}) \propto \mathbb{P}(\mathbf{x}|\theta = 0.6) \mathbb{P}(\theta = 0.6) = 0.02304 \times 0.94 = 0.021658$$

So $\hat{\theta} = 0.6$. Note the effect of this prior — by setting $\mathbb{P}(\theta = 0.6)$ so much higher than the other two values, we actually get a different maximizer. This is the effect of the prior on the MAP estimate.

4. You can simply choose $\mathbb{P}(\theta) = 1$ for $\theta = 0.3$, and zero for the other two cases.
5. They converge! The influence of the prior is significant when our samples are limited; however, as the number of samples increases, the evidence overwhelms the prior. Consider if we only toss the coin five times—the prior would have a substantial effect on our estimation. But if we toss the coin 10,000 times, the impact of any (small) prior would become negligible in our estimate.

Question 11

Speckle Generation: When a laser beam reflects off a rough surface, the scattered waves interfere and form a random pattern called ‘speckle’. The intensity distribution of the speckle pattern can be modeled probabilistically.

1. The electric field S of the scattered wave at a given point can be represented as a sum of N random ‘phasors’: $S = \sum_{n=1}^N A_n e^{i\phi_n}$, where A_n and ϕ_n are random amplitude and random phase for the n -th phasor. Assume that A_n and ϕ_n are independent of each other. Moreover, $\phi_n \sim U(-\pi, \pi)$ and A_n are i.i.d. for $n = 1, \dots, N$.

Let $S = X + iY$, where X and Y are the real and imaginary parts of the complex sum.

- (a) Find the expressions for X and Y in terms of the random variables A_n and ϕ_n .
 - (b) Show that for a very large N , both X and Y become approximately Gaussian distributed with $X, Y \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \frac{N}{2} \mathbb{E}(A_n^2)$.
2. Let the intensity of the speckle be the magnitude of the complex sum: $R = |S| = \sqrt{X^2 + Y^2}$.
 - (a) Use the joint distribution of X and Y in polar coordinates to derive the PDF of the speckle intensity R .
 - (b) Derive the CDF, mean, median, and mode for the intensity of the speckle pattern.
 3. Suppose that you are given a set of M independent intensity measurements $R_1, R_2, \dots, R_M$ for the speckle pattern.
 - (a) Determine the Maximum Likelihood estimator to estimate σ and σ^2 .
 - (b) Determine whether the estimators are unbiased.

See the answer on the next page.

1. (a) The complex field is given by $S = \sum_{n=1}^N A_n e^{i\phi_n}$. We can expand the exponential term using Euler's formula to get $S = \sum_{n=1}^N A_n (\cos(\phi_n) + i \sin(\phi_n))$.

Separating the real and imaginary parts, we have

$$S = \left(\sum_{n=1}^N A_n \cos(\phi_n) \right) + i \left(\sum_{n=1}^N A_n \sin(\phi_n) \right).$$

Therefore,

$$X = \sum_{n=1}^N A_n \cos(\phi_n) \quad \text{and} \quad Y = \sum_{n=1}^N A_n \sin(\phi_n).$$

- (b) For large N , by the central limit theorem, the sums of the cosine and sine components tend toward a Gaussian distribution by the aggregation of many small random contributions.

- i. The mean of X is:

$$\mathbb{E}(X) = \mathbb{E} \left(\sum_{n=1}^N A_n \cos(\phi_n) \right) = \sum_{n=1}^N \mathbb{E}(A_n \cos(\phi_n)) = \sum_{n=1}^N \mathbb{E}(A_n) \cdot \mathbb{E}(\cos(\phi_n)).$$

For uniformly distributed random phases $\phi_n \sim U(-\pi, \pi)$, we have

$$\mathbb{E}(\cos(\phi_n)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(\phi_n) d\phi_n = 0.$$

Therefore, $\mathbb{E}(X) = 0$, and similarly, $\mathbb{E}(Y) = 0$.

- ii. To calculate the variance of X , since the phases ϕ_n are uniformly distributed,

$$\mathbb{E}(\cos^2(\phi_n)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos^2(\phi_n) d\phi_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 + \cos(2\phi_n)}{2} d\phi_n = \frac{1}{2}.$$

By the independence property, for the random variable X , we can obtain

$$\begin{aligned} \text{var}(X) &= \text{var} \left(\sum_{n=1}^N A_n \cos(\phi_n) \right) = \sum_{n=1}^N \text{var}(A_n \cos(\phi_n)) \\ &= \sum_{n=1}^N \mathbb{E}(A_n^2 \cos^2(\phi_n)) - [\mathbb{E}(A_n \cos(\phi_n))]^2 \\ &= \sum_{n=1}^N \mathbb{E}(A_n^2) \mathbb{E}(\cos^2(\phi_n)) - [\mathbb{E}(A_n) \mathbb{E}(\cos(\phi_n))]^2 \\ &= \sum_{n=1}^N \mathbb{E}(A_n^2) \cdot \frac{1}{2} - [\mathbb{E}(A_n) \cdot 0]^2 = \frac{N}{2} \mathbb{E}(A_n^2). \end{aligned}$$

The calculation for $\text{var}(Y)$ is similar. Therefore, both of them are zero-mean Gaussian random variable $\mathcal{N}(0, \sigma^2)$, with $\sigma^2 = \frac{N}{2} \mathbb{E}(A_n^2)$.

2. (a) The joint PDF of X and Y (since they are i.i.d. Gaussian) is:

$$f_{X,Y}(x,y) = \frac{1}{\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)}.$$

Transform the Cartesian coordinates to polar coordinates:

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

The Jacobian of the transformation is:

$$dx dy = r dr d\theta.$$

The joint PDF in polar form is:

$$f_{R,\Theta}(r,\theta) = \frac{1}{\pi\sigma^2} e^{-r^2/(2\sigma^2)} \cdot r.$$

To get the marginal PDF of R , integrate over θ :

$$f_R(r) = \int_{-\pi}^{\pi} f_{R,\Theta}(r,\theta) d\theta, \quad r \geq 0.$$

Since the PDF is independent of θ , the integration yields:

$$f_R(r) = \begin{cases} \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)} & r \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The detailed derivations can refer to lecture 4 slides on page 23-27.

2. (b)

i. The CDF:

$$F_X(x) = \int_0^x \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)} dr$$

Let $y = \frac{r^2}{2\sigma^2}$. Then, $dy = \frac{r dr}{\sigma^2}$. Moreover, when $r = 0$, we have $y = 0$; when $r = x$, we have $y = \frac{x^2}{2\sigma^2}$. Therefore,

$$F_X(x) = \int_0^{x^2/2\sigma^2} e^{-y} dy = [-e^{-y}]_0^{x^2/2\sigma^2} = 1 - e^{-x^2/(2\sigma^2)}$$

ii. The mean:

$$\begin{aligned} \mathbb{E}(X) &= \int_0^\infty (1 - F_X(x)) dx = \int_0^\infty e^{-x^2/(2\sigma^2)} dx \\ &= \sigma\sqrt{\pi} \int_0^\infty \frac{1}{\sqrt{\pi}\sigma} e^{-x^2/(2\sigma^2)} dx \\ &= \sigma\sqrt{\pi} \cdot \frac{1}{2} \\ &= \sigma\sqrt{\frac{\pi}{2}} \end{aligned}$$

Because $\int_0^\infty \frac{1}{\sqrt{\pi}\sigma} e^{-x^2/(2\sigma^2)} dx$ is the area under the positive side of a zero-mean Gaussian.

iii. The median:

$$\begin{aligned} 1 - e^{-x^2/(2\sigma^2)} &= \frac{1}{2} \\ -x^2/(2\sigma^2) &= \ln \frac{1}{2} \\ x^2 &= 2\sigma^2 \ln 2 \\ x &= \sigma\sqrt{2 \ln 2} \end{aligned}$$

iv. The mode:

We find the maximum of $f_X(x)$ by taking the derivative with respect to x and setting it to zero:

$$\frac{d}{dx} \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} = \frac{1}{\sigma^2} e^{-x^2/(2\sigma^2)} - \frac{x}{\sigma^2} \cdot \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} = 0$$

This gives $x^2 = \sigma^2$, and therefore, $x = \sigma$.

3. (a) Given a set of independent intensity measurements $R_1, R_2, \dots, R_M$, the likelihood function $L(\sigma)$ is the product of the individual PDFs:

$$L(\sigma) = \prod_{m=1}^M f_R(r_m; \sigma) = \prod_{m=1}^M \frac{r_m}{\sigma^2} e^{-r_m^2/(2\sigma^2)}.$$

To simplify, take the logarithm of the likelihood function:

$$\log L(\sigma) = \sum_{m=1}^M \left(\log r_m - 2 \log \sigma - \frac{r_m^2}{2\sigma^2} \right).$$

Differentiate the log-likelihood function with respect to σ and set it equal to zero:

$$\frac{d}{d\sigma} \log L(\sigma) = -\frac{2M}{\sigma} + \frac{1}{\sigma^3} \sum_{m=1}^M r_m^2 = 0.$$

Solving for σ , we have

$$\hat{\sigma} = \sqrt{\frac{1}{2M} \sum_{m=1}^M R_m^2} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{2M} \sum_{m=1}^M R_m^2.$$

- (b) Compute the expectation of $\hat{\sigma}^2$:

$$\begin{aligned} \mathbb{E}(\hat{\sigma}^2) &= \mathbb{E}\left(\frac{1}{2M} \sum_{m=1}^M R_m^2\right) = \frac{1}{2M} \sum_{m=1}^M \mathbb{E}(R_m^2) \\ &= \frac{1}{2M} \sum_{m=1}^M \mathbb{E}(X_m^2 + Y_m^2) \\ &= \frac{1}{2M} \sum_{m=1}^M 2\sigma^2 = \sigma^2. \end{aligned}$$

So, $\hat{\sigma}^2$ is an unbiased estimator for σ^2 .

On the other hand, when we compute the expectation of $\hat{\sigma}$:

$$\mathbb{E}(\hat{\sigma}) = \mathbb{E}\left(\sqrt{\frac{1}{2M} \sum_{m=1}^M R_m^2}\right) < \sqrt{\mathbb{E}\left(\frac{1}{2M} \sum_{m=1}^M R_m^2\right)} = \sigma.$$

A square root is a nonlinear transformation. From Jensen's inequality, $\mathbb{E}(\sqrt{A}) \leq \sqrt{\mathbb{E}(A)}$, with equality if and only if A is a constant, which is not satisfied in this problem. Therefore, $\hat{\sigma}$ will tend to underestimate σ , making it a biased estimator. (Note: since the question asks whether it is an unbiased estimator, as long as you can show that $\mathbb{E}(\hat{\sigma}) \neq \sigma$, that would be sufficient. You are not required to invoke Jensen's inequality.)

Question 12

In a game, the players need to control a game character to defeat as many monsters as possible. The players will only encounter one monster each round. A player's chance of defeating a monster in each round is p . In addition, the game has the following characteristics:

- The game character has unlimited number of opportunities to defeat the monsters, although in each round, there is still only one monster.
- Once a player has defeated r number of monsters, he or she is considered to have "won the game."
- Let K denote the number of failures before winning the game. This is a discrete random variable.

Our goal for this problem is to estimate p , based on the observed values of K for many, many players. We denote these observations as $k_1, k_2, \dots, k_n$.

1. What is the probability mass function of K ?
2. What is the maximum likelihood estimate of p ?
3. Assume that p follows a beta distribution (i.e., the prior), with

$$f_p(p) \sim \text{Beta}(p; \alpha, \beta) = \begin{cases} \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} & 0 < p < 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $B(\alpha, \beta)$ is a normalizing constant, known as the beta function, given by

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \frac{(\alpha - 1)!(\beta - 1)!}{(\alpha + \beta - 1)!}$$

What is the MAP estimate of p ?

See the answer on the next page.

1. The probability of $K = k$ can be derived as follows. In the first $r + k - 1$ rounds, the player should win $r - 1$ times and fail k times. Then, in the last round, the player wins. This probability is

$$\begin{aligned} p_K(k) &= \binom{r+k-1}{r-1} (1-p)^k (p)^{r-1} \cdot (p) \\ &= \binom{r+k-1}{r-1} (1-p)^k (p)^r \end{aligned}$$

2. The likelihood function is

$$\begin{aligned} p_K(k_1, \dots, k_n; p) &= \prod_{i=1}^n \binom{r+k_i-1}{r-1} (1-p)^{k_i} (p)^r \\ &\propto (1-p)^{\sum_{i=1}^n k_i} (p)^{nr} \end{aligned}$$

The log-likelihood function is

$$\log p_K(k_1, \dots, k_n; p) = C + \sum_{i=1}^n k_i \log(1-p) + nr \log p$$

where C is a certain constant. Taking the derivative with respect to p and setting it to zero, we have

$$\frac{nr}{p} = \frac{\sum_{i=1}^n k_i}{1-p}$$

So the ML estimator is

$$\hat{p}_{n,ML} = \frac{nr}{\sum_{i=1}^n k_i + nr}$$

3. The posterior is the product of the likelihood and the prior, i.e.,

$$\begin{aligned} f_{P|K_1, \dots, K_n}(p | k_1, \dots, k_n) &\propto (1-p)^{\sum_{i=1}^n k_i} (p)^{nr} \cdot p^{\alpha-1} (1-p)^{\beta-1} \\ &= (1-p)^{\sum_{i=1}^n k_i + \beta - 1} (p)^{nr + \alpha - 1} \end{aligned}$$

Following a similar procedure as above, by taking the derivative with respect to p and setting it to zero, we can find the MAP estimator to be

$$\hat{p}_{n,MAP} = \frac{nr + \alpha - 1}{\sum_{i=1}^n k_i + nr + \alpha + \beta - 2}$$

Question 13

The number of photons X arriving at a detector within the detection duration t follows a Poisson distribution as

$$\mathbb{P}(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad k = 0, 1, \dots$$

where the parameter λ indicates the light intensity. In this question, we will explore two different ways of obtaining the light intensity λ via recoding and timing the arrival photons.

1. In the first way, the number of arrival photons within the detection duration t is encoded as a binary signal Y , with 0 indicating no photon arriving at the detector and 1 indicating there is at least one arrival photon.
 - (a) Show that Y follows a Bernoulli distribution, *i.e.*, $Y \sim \text{Ber}(\theta)$, where the parameter θ is determined by λ .
 - (b) Given a sequence of independent and identically distributed observations $y_1, \dots, y_n$ of the random variable Y , find the ML estimate of λ . (Hint: You can find the ML estimate of θ and determine λ using the relationship between θ and λ .)
2. In the second way, the detection duration is assumed to be long enough so that there are more than two arrival photons, and the time interval Z between the consecutive arrival photons is recorded.
 - (a) Show that the time interval Z follows an exponential distribution of which λ is the parameter, *i.e.*, $Z \sim \text{Exp}(\lambda)$.
 - (b) Given a set of independent and identically distributed observations $z_1, \dots, z_n$ of the random variable Z , find the ML estimate of λ .

See the answer on the next page.

1. (a) For the binary random variable Y , we have

$$\begin{aligned}\mathbb{P}(Y = 0) &= \mathbb{P}(X = 0) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}, \\ \mathbb{P}(Y = 1) &= \mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-\lambda t}.\end{aligned}$$

Therefore, Y follows the Bernoulli distribution $\text{Ber}(\theta)$ of which the parameter is

$$\theta = \mathbb{P}(Y = 1) = 1 - e^{-\lambda t}.$$

- (b) Given independent and identically distributed observations $y_1, \dots, y_n$ of the random variable $Y \sim \text{Ber}(\theta)$, the likelihood can be written as

$$p_Y(y_1, \dots, y_n; \theta) = \prod_{i=1}^n (1 - \theta)^{1-y_i} \theta^{y_i} = (1 - \theta)^{n-m} \theta^m,$$

where $m = \sum_{i=1}^n y_i$. The log-likelihood is then

$$\log p_Y(y_1, \dots, y_n; \theta) = (n - m) \log(1 - \theta) + m \log \theta.$$

Now we can take the derivative (w.r.t. θ) and set it to zero, such that

$$-\frac{n - m}{1 - \theta} + \frac{m}{\theta} = 0.$$

Thus, the solution of this equation is $\theta_{\text{ML}} = \frac{m}{n}$. Furthermore, using the relation between θ and λ derived above, which is that $\theta = 1 - e^{-\lambda t}$, we can obtain the ML estimate of λ as

$$\lambda_{\text{ML}} = -\log(1 - \theta_{\text{ML}}) = -\log\left(1 - \frac{m}{n}\right) = -\log\left(1 - \frac{1}{n} \sum_{i=1}^n y_i\right).$$

2. (a) According to the connection between exponential and Poisson random variables, as introduced in page 36 of Lecture 3, we can directly conclude that the time interval Z follows $\text{Exp}(\lambda)$ when the number of arrival photons X follows $\text{Poisson}(\lambda t)$. The detailed proof is also provided as follows:

First, we note that

$$\mathbb{P}(Z > t) = \mathbb{P}(\text{interval time} > t) = \mathbb{P}(\text{no arrival within } t) = \mathbb{P}(X = 0) = e^{-\lambda t}.$$

Then, the CDF of Z can be obtained as

$$F_Z(t) = 1 - \mathbb{P}(Z > t) = 1 - e^{-\lambda t}.$$

Thus, $Z \sim \text{Exp}(\lambda)$.

- (b) Given independent and identically distributed observations $z_1, \dots, z_n$ of the random variable $Z \sim \text{Exp}(\lambda)$, the likelihood can be written as

$$f_Z(z_1, \dots, z_n; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda z_i} = \lambda^n e^{-\lambda T},$$

where $T = \sum_{i=1}^n z_i$. The log-likelihood is then

$$\log f_Z(z_1, \dots, z_n; \lambda) = n \log \lambda - \lambda T.$$

Now we can take the derivative (w.r.t. λ) and set it to zero, such that

$$\frac{n}{\lambda} - T = 0.$$

Thus, the solution of this equation is $\lambda_{\text{ML}} = \frac{n}{T}$.

Question 14

In cell biology, protein-protein interactions (PPIs) refer to contacts established between two protein molecules due to some biochemical events. They are foundational to carrying out various biological functions, such as DNA replication. In a collection of proteins, such interactions form a protein-protein interaction network (PPIN).

In this problem, we want to study a PPIN. Suppose certain drugs would facilitate the development of an interaction between two proteins. However, whether a PPI is formed between them is a random process, modeled by a Bernoulli random variable. Specifically, for protein P_i and protein P_j , where $i \neq j$ (a protein would not form a PPI with itself), their PPI is denoted as X_{ij} where

$$X_{ij} = \begin{cases} 1 & \text{if } (P_i, P_j) \text{ have interactions} \\ 0 & \text{otherwise.} \end{cases}$$

Assume, for simplicity, that they all follow the same distribution, i.e.,

$$X_{ij} \sim \text{Ber}(q) \quad \text{for } i \neq j.$$

The parameter q can be thought of as the *effectiveness* of the drug to form PPIs. Our goal is to estimate this q using maximum likelihood (ML).

1. What is the PMF of X_{ij} , denoted as $p_{X_{ij}}(x_{ij})$?
2. In a PPIN with N proteins, we observed the various x_{ij} . What is the log likelihood, denoted as $\log p_X(\mathbf{x}; q)$? (*note: X here would denote the list of all X_{ij} .*)
3. What is the ML estimate of q ?

1. As X_{ij} follows a Bernoulli distribution, we can write the PMF as

$$p_{X_{ij}}(x_{ij}) = \begin{cases} q & x_{ij} = 1 \\ 1 - q & x_{ij} = 0 \end{cases}$$

This can be combined as a single equation, i.e.,

$$p_{X_{ij}}(x_{ij}) = q^{x_{ij}} (1 - q)^{1-x_{ij}}$$

2. The log likelihood is given by

$$\begin{aligned} \log p_X(\mathbf{x}; q) &= \log \prod_{i=1}^N \prod_{\substack{j=1 \\ i \neq j}}^N q^{x_{ij}} (1 - q)^{1-x_{ij}} \\ &= \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N (x_{ij} \log q + (1 - x_{ij}) \log(1 - q)) \end{aligned}$$

3. Taking the derivative with respect to q and setting it to zero, we have

$$\begin{aligned} \sum \sum \left(\frac{x_{ij}}{q} - \frac{1 - x_{ij}}{1 - q} \right) &= 0 \\ \frac{\sum \sum x_{ij}}{q} &= \frac{k - \sum \sum x_{ij}}{1 - q} \\ q &= \frac{1}{k} \sum \sum x_{ij} \end{aligned}$$

where $k = N(N - 1)$, the total number of possible connections in the PPIN, and the summation is for all values of i, j except $i = j$ (same as what's written in the previous part).

Question 15

Suppose we collect n independent observations $Y_1, Y_2, \dots, Y_n$, each of which follows a normal distribution with mean equal to the product of a single unknown parameter θ and known constants $x_1, x_2, \dots, x_n$, and with a known variance σ^2 . In other words, $Y_i \sim \mathcal{N}(\theta x_i, \sigma^2)$, where x_i and σ^2 are fixed and known, while θ must be estimated.

1. Find the maximum likelihood estimator $\hat{\theta}$.
2. Before seeing any data, we assume a normal distribution prior on θ , with specified mean equal to θ_0 and variance equal to τ^2 .
 - (a) Derive the posterior distribution of θ . (Hint: It is also a normal distribution, but you need to derive expressions for the mean (θ_1) and variance (σ_1^2) of this normal distribution.)
 - (b) Show that we can express the mean of the posterior distribution as a weighted sum of the prior mean θ_0 and the maximum likelihood estimator $\hat{\theta}$.
 - (c) Find the limit of the posterior distribution as $\tau^2 \rightarrow \infty$, and compare the estimates given by ML and MAP.

1. We have $Y_i \sim \mathcal{N}(\theta x_i, \sigma^2)$. Therefore,

$$f_{Y|\Theta}(y_1, \dots, y_n | \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(y_i - \theta x_i)^2}{2\sigma^2} \right\}$$

$$\log(f_{Y|\Theta}(y_1, \dots, y_n | \theta)) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta x_i)^2$$

$$\frac{d}{d\theta} \log(f_{Y|\Theta}(y_1, \dots, y_n | \theta)) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \theta x_i) x_i$$

Setting the above expression to zero, we obtain

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

- 2 (a) The prior is $f_{\Theta}(\theta) = \frac{1}{\sqrt{2\pi\tau}} \exp \left\{ -\frac{(\theta - \theta_0)^2}{2\tau^2} \right\}$. Therefore, the posterior is

$$\begin{aligned} f_{\Theta|Y}(\theta|y_1, \dots, y_n) &\propto f_{Y|\Theta}(y_1, \dots, y_n|\theta) f_{\Theta}(\theta) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta x_i)^2 - \frac{1}{2\tau^2} (\theta - \theta_0)^2 \right\} \\ &= \exp \left\{ -\frac{1}{2} \underbrace{\left[\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \theta x_i)^2 + \frac{1}{\tau^2} (\theta - \theta_0)^2 \right]}_{=M} \right\} \end{aligned}$$

Since we know that the posterior is a normal distribution, our goal is to derive θ_1 and σ_1 , such that

$$M = \frac{(\theta - \theta_1)^2}{\sigma_1^2} + K,$$

where K is a certain constant. We proceed by expanding the terms of M and then re-grouping them. Henceforth, all summations are from $i = 1$ to n .

$$\begin{aligned} M &= \frac{\sum y_i^2}{\sigma^2} - \frac{2\theta \sum x_i y_i}{\sigma^2} + \frac{\theta^2 \sum x_i^2}{\sigma^2} + \frac{\theta^2}{\tau^2} - \frac{2\theta\theta_0}{\tau^2} + \frac{\theta_0^2}{\tau^2} \\ \frac{\theta^2}{\sigma_1^2} - 2\theta \frac{\theta_1}{\sigma_1^2} + \frac{\theta_1^2}{\sigma_1^2} + K &= \theta^2 \left(\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2} \right) - 2\theta \left(\frac{\sum x_i y_i}{\sigma^2} + \frac{\theta_0}{\tau^2} \right) + \left(\frac{\sum y_i^2}{\sigma^2} + \frac{\theta_0^2}{\tau^2} \right) \end{aligned}$$

Comparing the two sides, we have

$$\begin{aligned} \sigma_1^2 &\triangleq \left(\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} \\ \theta_1 &\triangleq \sigma_1^2 \left(\frac{\sum x_i y_i}{\sigma^2} + \frac{\theta_0}{\tau^2} \right) \end{aligned}$$

- (b) We can express θ_1 as

$$\begin{aligned} \theta_1 &= \frac{\frac{\sum x_i y_i}{\sigma^2} + \frac{\theta_0}{\tau^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} = \frac{\frac{\sum x_i y_i}{\sum x_i^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2 \sum x_i^2}} + \frac{\theta_0}{\tau^2 \left(\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2} \right)} \\ &= \left[\frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2 \sum x_i^2}} \right] \hat{\theta} + \left[\frac{1}{\tau^2 \left(\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2} \right)} \right] \theta_0 \end{aligned}$$

- (c) As $\tau^2 \rightarrow \infty$, $\theta_1 \rightarrow \hat{\theta}$. Therefore, the MAP estimate approaches the ML estimate.