

Homework 3

Due: Apr 30, 2026, 11:55pm

Instruction: Submit your answers electronically through Moodle. In Moodle, you must submit under the PDF Submission link with ONE PDF file containing answers to all remaining questions.

Question 1

A joint PDF of random variables X and Y is given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x-\mu_x)}{\sigma_x} \frac{(y-\mu_y)}{\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right]},$$

where ρ denotes the correlation, with $-1 < \rho < 1$.

1. Show that if $\rho = 0$, then X and Y are independent, with each following a Gaussian PDF.
2. Show that for a general ρ but with $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$:

(a) The expression can be simplified to

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

(b) Let $E = \frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}$. Show that we can also write

$$E = \frac{(y - \rho x)^2}{2(1-\rho^2)} + \frac{x^2}{2}.$$

(c) Let $u = y - \rho x$, and $\sigma^2 = 1 - \rho^2$. Show that the marginal PDF, given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy,$$

is also a Gaussian PDF.

1. If $\rho = 0$, then

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2} \end{aligned}$$

So we can see that X and Y are independent, with each following a Gaussian PDF.

2. (a) We can just plug in $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$ and the result follows directly.

(b)

$$E = \frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} = \frac{y^2 - 2\rho xy + \rho^2 x^2 + x^2 - \rho^2 x^2}{2(1-\rho^2)} = \frac{(y - \rho x)^2}{2(1-\rho^2)} + \frac{x^2}{2}.$$

- (c) We note that $du = dy$, and that when $y = \pm\infty$, we also have $u = \pm\infty$. Therefore,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\left(\frac{u^2}{2\sigma^2} + \frac{x^2}{2}\right)} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{u^2}{2\sigma^2}\right)} du \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \end{aligned}$$

because the term under the last integration is a zero-mean Gaussian, which has an area of 1. The expression for $f_X(x)$ is a zero-mean Gaussian with unit variance.

Question 2

A certain continuous random variable X has probability density function (PDF) given by

$$f_X(x) = \begin{cases} \frac{k}{\lambda^k} x^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

There are two parameters:

- A scale parameter, λ , which must be positive.
- A shape parameter, k , which must also be positive.

We made observations $X_1, \dots, X_n$, which are independent identically distributed and follow this PDF.

1. Suppose $k = 1$. Find the ML estimator of λ .
2. Suppose $k = 2$. Find the ML estimator of λ .
3. Find the ML estimator of λ for a general value of k .

1. When $k = 1$, the PDF becomes

$$f_X(x) = \begin{cases} \frac{1}{\lambda} e^{-(x/\lambda)} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

This is an exponential distribution, with parameter $\frac{1}{\lambda}$.

The log-likelihood function is given by

$$\log f_X(x_1, \dots, x_n; \lambda) = \sum_{i=1}^n \log \frac{1}{\lambda} e^{-x_i/\lambda} = -n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^n x_i$$

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

$$-\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^n x_i = 0$$

$$\lambda = \frac{1}{n} \sum_{i=1}^n x_i$$

Hence, the ML estimate is

$$\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

2. When $k = 2$, the PDF becomes

$$f_X(x) = \begin{cases} \frac{2x}{\lambda^2} e^{-(x^2/\lambda^2)} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

This is a Rayleigh distribution, with parameter $\frac{\lambda}{\sqrt{2}}$.

The log-likelihood function is given by

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

Now, we take derivative with respect to λ , and set to zero:

$$\begin{aligned} \frac{d \log f_X(x_1, \dots, x_n; \lambda)}{d\lambda} &= \frac{d}{d\lambda} \sum_{i=1}^n \left(\log x_i + \log 2 - 2 \log \lambda - \frac{x_i^2}{\lambda^2} \right) \\ &= \sum_{i=1}^n \left(-\frac{2}{\lambda} + \frac{2x_i^2}{\lambda^3} \right) \\ &= -\frac{2}{\lambda} \sum_{i=1}^n \left(1 - \frac{x_i^2}{\lambda^2} \right) = 0 \end{aligned}$$

This gives $\lambda^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$. Hence, the ML estimate is

$$\hat{\lambda}_n = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

3. For a general k , the ML estimate is

$$\hat{\lambda}_n = \left(\frac{1}{n} \sum_{i=1}^n x_i^k \right)^{1/k}$$

following similar derivations as above.

Question 3

In an autonomous driving system, a radar module measures the distance to a moving vehicle. Suppose we collect n independent sensor readings $Y_1, Y_2, \dots, Y_n$. Each reading follows a normal distribution where the mean is proportional to an unknown velocity parameter θ and known time instances $x_1, x_2, \dots, x_n$. The sensor noise variance σ^2 is known. 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 sensor data, the car's predictive tracking algorithm assumes 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. The likelihood function and its log-likelihood are:

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

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

Taking the derivative and setting it to zero:

$$\frac{d}{d\theta} \log f_{Y|\Theta}(y_1, \dots, y_n | \theta) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \theta x_i) x_i = 0 \implies \hat{\theta}_{ML} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

2. For simplicity in the mathematical expressions, the limits of the summation signs will be omitted. They are taken to be from $i = 1 \dots n$ unless stated otherwise.

(a) The posterior distribution is proportional to the likelihood multiplied by the prior:

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

As suggested by the hint, we want to show that this posterior is also a normal distribution. That means we can write

$$\frac{\sum (y_i - \theta x_i)^2}{\sigma^2} + \frac{(\theta - \theta_0)^2}{\tau^2} = \frac{(\theta - \theta_1)^2}{\sigma_1^2} + k$$

where k is a certain constant. All terms that do not contain θ will end up contributing to the normalization.

Expanding the terms on both sides, we have

$$\frac{\sum y_i^2}{\sigma^2} - 2\theta \frac{\sum x_i y_i}{\sigma^2} + \theta^2 \frac{\sum x_i^2}{\sigma^2} + \frac{\theta_0^2}{\tau^2} - 2\theta \frac{\theta_0}{\tau^2} + \theta^2 \frac{1}{\tau^2} = \frac{\theta_1^2}{\sigma_1^2} - 2\theta \frac{\theta_1}{\sigma_1^2} + \theta^2 \frac{1}{\sigma_1^2} + k$$

Equating the coefficients for θ and θ^2 , we obtain:

$$\sigma_1^2 = \left(\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1}, \quad \theta_1 = \sigma_1^2 \left(\frac{\sum x_i y_i}{\sigma^2} + \frac{\theta_0}{\tau^2} \right)$$

- (b) Since the posterior is a normal distribution, the mean is θ_1 . Substituting the expression for σ_1^2 into θ_1 , we obtain

$$\theta_1 = \frac{\frac{\sum x_i y_i}{\sigma^2} + \frac{\theta_0}{\tau^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} = \left[\frac{\frac{\sum x_i^2}{\sigma^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right] \hat{\theta}_{ML} + \left[\frac{\frac{1}{\tau^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right] \theta_0$$

This clearly shows that the MAP estimate (θ_1) is a weighted average of the ML estimate and the prior mean, where the two weights sum to exactly 1.

- (c) As $\tau^2 \rightarrow \infty$, the prior becomes completely uninformative. The term $\frac{1}{\tau^2} \rightarrow 0$. Therefore, the weight on θ_0 goes to 0, and the weight on $\hat{\theta}_{ML}$ goes to 1.

$$\lim_{\tau^2 \rightarrow \infty} \theta_1 = \hat{\theta}_{ML}$$

This demonstrates that with an uninformative prior, the MAP estimate approaches the ML estimate.

Question 4

A mobile application records the type of action performed by users when they open the app.

Each user action falls into one of three categories:

Browsing, Messaging, Shopping.

Let the probabilities of the three actions be θ_1 , θ_2 , and θ_3 , respectively, and assume that none of them is zero. Note also that $\theta_1 + \theta_2 + \theta_3 = 1$.

Before observing any data, an engineering intern, equipped with knowledge of probabilistic systems analysis, models the prior belief of the parameters by a Dirichlet distribution

$$(\theta_1, \theta_2, \theta_3) \sim \text{Dir}(\alpha_1, \alpha_2, \alpha_3).$$

The app then records n user sessions independently. Suppose that the observations show k_1 browsing actions, k_2 messaging actions, and k_3 shopping actions, with $k_1 + k_2 + k_3 = n$.

1. Write down the likelihood function of $(\theta_1, \theta_2, \theta_3)$.
2. Find the maximum likelihood estimators for $\theta_1, \theta_2, \theta_3$.
3. Derive the posterior distribution of $(\theta_1, \theta_2, \theta_3)$ and find the MAP estimators.

1. Since each observation falls into one of the three categories, the counts (k_1, k_2, k_3) follow a multinomial model.

Hence the likelihood function is

$$L(\theta_1, \theta_2, \theta_3) = \frac{n!}{k_1! k_2! k_3!} \theta_1^{k_1} \theta_2^{k_2} \theta_3^{k_3}, \quad \text{with } k_1 + k_2 + k_3 = n.$$

2. To find the maximum likelihood estimators, we maximize $L(\theta_1, \theta_2, \theta_3)$ subject to $\theta_1 + \theta_2 + \theta_3 = 1$.

Since the multinomial coefficient does not depend on the parameters, it is enough to maximize the log-likelihood

$$\log L(\theta_1, \theta_2, \theta_3) = C + k_1 \log \theta_1 + k_2 \log \theta_2 + k_3 \log \theta_3,$$

where C is a certain constant.

We can take care of the constraint by making the substitution $\theta_3 = 1 - \theta_1 - \theta_2$, such that

$$\log L(\theta_1, \theta_2, \theta_3) = C + k_1 \log \theta_1 + k_2 \log \theta_2 + k_3 \log(1 - \theta_1 - \theta_2).$$

Taking partial derivatives and setting them equal to zero, we have

$$\frac{k_1}{\theta_1} = \frac{k_3}{1 - \theta_1 - \theta_2}, \quad \frac{k_2}{\theta_2} = \frac{k_3}{1 - \theta_1 - \theta_2}.$$

Because $1 - \theta_1 - \theta_2 = \theta_3$, we have

$$\frac{\theta_1}{k_1} = \frac{\theta_2}{k_2} = \frac{\theta_3}{k_3}.$$

Hence, for some constant c ,

$$\theta_1 = ck_1, \quad \theta_2 = ck_2, \quad \theta_3 = ck_3.$$

Using $\theta_1 + \theta_2 + \theta_3 = 1$ and $k_1 + k_2 + k_3 = n$, we get $c = \frac{1}{n}$. Therefore,

$$\hat{\theta}_{1,ML} = \frac{k_1}{n}, \quad \hat{\theta}_{2,ML} = \frac{k_2}{n}, \quad \hat{\theta}_{3,ML} = \frac{k_3}{n}.$$

3. The prior distribution is $(\theta_1, \theta_2, \theta_3) \sim \text{Dir}(\alpha_1, \alpha_2, \alpha_3)$, so its density is

$$f_{\theta}(\theta_1, \theta_2, \theta_3) \propto \theta_1^{\alpha_1-1} \theta_2^{\alpha_2-1} \theta_3^{\alpha_3-1}.$$

From part (1), the likelihood is $L(\theta_1, \theta_2, \theta_3) \propto \theta_1^{k_1} \theta_2^{k_2} \theta_3^{k_3}$. Hence,

$$\begin{aligned} f_{\theta}(\theta_1, \theta_2, \theta_3 | X) &\propto f_{\theta}(\theta_1, \theta_2, \theta_3) L(\theta_1, \theta_2, \theta_3) \\ &\propto \theta_1^{\alpha_1+k_1-1} \theta_2^{\alpha_2+k_2-1} \theta_3^{\alpha_3+k_3-1}. \end{aligned}$$

Therefore,

$$(\theta_1, \theta_2, \theta_3) | X \sim \text{Dir}(\alpha_1 + k_1, \alpha_2 + k_2, \alpha_3 + k_3).$$

From the formula given in Lecture 6 for the MAP estimate of a Dirichlet posterior,

$$\theta_k = \frac{\alpha_k + x_k - 1}{\alpha_0 + n - K} \quad \text{with } \alpha_0 = \alpha_1 + \alpha_2 + \alpha_3.$$

Here $K = 3$, and the observed counts are $x_k = k_k$ for $k = 1, 2, 3$. Hence

$$\begin{aligned} \hat{\theta}_{1,MAP} &= \frac{\alpha_1 + k_1 - 1}{\alpha_1 + \alpha_2 + \alpha_3 + n - 3} \\ \hat{\theta}_{2,MAP} &= \frac{\alpha_2 + k_2 - 1}{\alpha_1 + \alpha_2 + \alpha_3 + n - 3} \\ \hat{\theta}_{3,MAP} &= \frac{\alpha_3 + k_3 - 1}{\alpha_1 + \alpha_2 + \alpha_3 + n - 3}. \end{aligned}$$

Question 5

A certain random variable X is modeled as an exponential distribution, with $X \sim \text{Exp}(\theta)$, i.e.,

$$f_X(x) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

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 exponential distribution.

1. Show that the posterior is given by

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

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

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

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

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

Therefore,

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

where c is a certain constant.

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

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

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

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

4. We can write

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

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

$$\hat{\theta} \approx \frac{n}{s} = \frac{1}{\frac{1}{n} \sum_{i=1}^N X_i}.$$

This is the ML estimate!