EECS 965 Exam II

Fall 2025 - 75 Points

Name William Powers

Instructions

1) Your solutions must be detailed, unambiguous, and organized. I require that you show **all** your mathematics (i.e., no Matlab unless explicitly allowed).

2) Mathematically reduce your solutions to their **simplest** and most fundamental form.

2) Please **underline or circle** all numeric answers (e.g., |**A**| = 11.2), or in some way make your final answer **clear**.

4) If you feel that a problem is unclear, contradictory, incomplete, or ambiguous, **ask for clarification**.

5) This is an exam; it must reflect **your** knowledge and effort—and yours only!

Please **sign** this statement: "*As an honorable scholar and human being, I pledge that this exam is a reflection of my knowledge only. I hereby pledge that I committed no act that a reasonable person could construe as academic misconduct.*"

Signed: W. Powers 12/11/2025

Problem 1 - 25 points

A scalar random variable x is **Gaussian** distributed as:

$$x \sim \mathcal{N}\left(0, \sigma^2\right)$$

a) Say we wish to estimate the **standard deviation** $\sigma = \theta$ from an observation x.

Determine the **Maximum Likelihood** estimator $\hat{\sigma}_{mle}(x)$

b) Say instead we wish to estimate the **variance** $\sigma^2 = \theta$ from an observation x .

Determine the **Maximum Likelihood** estimator $\widehat{\sigma^2}_{mle}(x)$

c) How is the **ML** estimate $\widehat{\sigma^2}_{mle}(x)$ mathematically related to the **ML** estimate $\hat{\sigma}_{mle}(x)$ (i.e., can one estimate be expressed in terms of the other?).

Does this result surprise you? **Explain** why or why not.

d) Use the **CRLB theorem** to determine whether $\widehat{\sigma^2}_{mle}(x)$ is (or is not) an **efficient** estimator.

e) Use the **CRLB theorem** to determine whether $\hat{\sigma}_{mle}(x)$ is (or is not) an **efficient** estimator.

f) Given the results of part **c),** do the results of **d)** and **e)** surprise you? **Explain** why or why not.

Hint: Be sure to use the CRLB **THEOREM** in parts **d)** and **e).** In other words, do **NOT** attempt to determine the CRLB directly!

(a) Say we wish to estimate the standard deviation $\sigma = \theta$ from observation x. Determine the Maximum Likelihood estimator $\widehat{\sigma}_{MLE}(x)$.

In classical estimation theory, the Maximum Likelihood Estimator (MLE) is defined as the parameter value that maximizes the likelihood function. Once the likelihood function is obtained, the problem reduces to finding its stationary points and determining whether each is a maximum, minimum, or saddle point. Here, the scalar random variable x is assumed to be Gaussian distributed as $x \sim \mathcal{N}(0, \sigma^2)$ and therefore the likelihood function likewise takes the standard Gaussian form. Being that the parameter to be estimated is the standard deviation σ, it appears as the parameter in the conditional density notated as θ.

$$p(x|\theta = \sigma) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left[-\frac{x^2}{2\theta^2}\right]$$

To determine the stationary points of the above function, its derivative with respect to the conditional variable is taken. Before this differentation is preformed, the natural logarithm operator is applied as will not effect the location of the stationary points—being that this operator is monotonically increasing—and greatly simplifies the exponential through both power and product logarithm rules.

$$\ln\left(p(x|\theta)\right) = \ln(1) - \ln\left(\sqrt{2\pi\theta^2}\right) - \frac{x^2}{2\theta^2} = -\frac{1}{2}\ln(2\pi) - \ln(\theta) - \frac{x^2}{2\theta^2}$$

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = -\frac{1}{\theta} + \frac{x^2}{\theta^3}$$

Setting this derivative equal to zero and solving for the estimation parameter θ results in

$$-\frac{1}{\theta} + \frac{x^2}{\theta^3} = 0 \quad\longrightarrow\quad \theta^3(*) \quad\longrightarrow\quad -\theta^2 + x^2 = 0 \quad\longrightarrow\quad x^2 = \theta^2 \quad\longrightarrow\quad \theta = \pm x,$$

which, being that θ represents a standard deviation that is always greater than or equal to zero,

$$\theta = |x|.$$

The second derivative test is now used to classify this single stationary point as either a global minimum or maximum. Differentiating the first derivative of the log-likelihood ratio above with respect to θ yeilds

$$\frac{\partial^2 \ln(p(x|\theta))}{\partial^2\theta} = \frac{1}{\theta^2} - \frac{3x^2}{\theta^4}$$

which is then evaluated at the stationary point $\theta = |x|$

$$\frac{1}{|x|^2} - \frac{3x^2}{|x|^4} = \frac{1}{|x|^2} - \frac{3}{|x|^2} = -\frac{2}{|x|^2} \quad\longrightarrow\quad -\frac{2}{|x|^2} < 0 \;\; \forall \;\; x \neq 0.$$

This result is negative for all $x \neq 0$ and proves the function is concave across its domain, $\theta = |x|$ is a global maximum of the likelihood function, and is therefore the MLe estimate.

$$\hat{\sigma}_{MLE}(x) = |x|$$

(b) Say instead we wish to estimate the variance $\sigma^2 = \theta$ from observation x. Determine the Maximum Likelihood estimator $\widehat{\sigma^2}_{MLE}(x)$.

Building upon the theory provided in the previous reponse, estimating the variance σ^2 follows a similar sequence in which the likelihood function is formed and then extreme points are classified to find the global maximum. Again, notating the paramter to be estimated as θ, the standard Gaussian likelihood function is

$$p(x|\theta = \sigma^2) = \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right]$$

Applying the natural logarithm operator and differentiating yeilds

$$\ln\left(p(x|\theta)\right) = \ln(1) - \ln\left(\sqrt{2\pi\theta}\right) - \frac{x^2}{2\theta} = -\frac{1}{2}\ln(2\pi\theta) - \frac{x^2}{2\theta}$$

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}$$

and setting this result equal to zero reveals the stationary point

$$-\frac{1}{2\theta} + \frac{x^2}{2\theta^2} = 0 \longrightarrow \theta^2[\ldots] \longrightarrow -\theta + x^2 = 0 \longrightarrow \theta = x^2.$$

The second derivative, again with respect to θ, is taken and the stationary point is substituted in to classify this extrema as a maximum.

$$\frac{\partial^2 \ln(p(x|\theta))}{\partial^2\theta} = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3} \longrightarrow \theta = x^2 \longrightarrow \frac{1}{2x^4} - \frac{x^2}{x^6} = -\frac{1}{2x^4} < 0 \ \forall \ x \neq 0$$

From this conclusion it is proven that $\theta = x^2$ maximizes the likelihood function and therefore is the MLE.

$$\widehat{\sigma^2}_{MLE}(x) = x^2$$

(c) How is the ML estimate $\widehat{\sigma^2}_{MLE}(x)$ mathmatically related to the ML estimate $\widehat{\sigma}_{MLE}(x)$ (i.e. can one estimate be expressed in terms of the other)? Does this result suprise you? Explain why or why not.

The ML estimate of the standard deviation and the ML estimate of the variance are directly related through a square: the variance estimate is simply the square of the standard-deviation estimate. This relationship is expected, since variance is defined as the square of the standard deviation.

$$\widehat{\sigma^2}_{MLE}(x) = x^2 = (|x|)^2 = (\widehat{\sigma}_{MLE})^2$$

(d) Use the CRLB theorem to determine whether $\widehat{\sigma^2}_{MLE}(x)$ is (or is not) and efficient estimator.

The Cramer-Rao Lower Bound (CRLB) theorm, derived from the Fishers information of the underlying measurement, states that an efficient estimator exists if and only if the derivative of the log-likelihood function can be expressed in the form

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = f(\theta)(g(x) - \theta).$$

Here the likelihood function is seperated into the product of a function solely dependent on the parameter to be estimated θ and a function solely dependent on the measurement x. Furthermore, the resulting function $g(x)$ is equal to the efficient ML estimator and the inverse of function $f(\theta)$ is equal to its associated error. This being said, if through algebraic manipulation the derivative found in part (b) can be expressed in this form and the resulting $g(x)$ is equal to the derived ML estimator, then that estimator is efficient. Restating the first derivative with respect to θ below

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}$$

the term $2\theta^2$ can be factored out of both denomonators resulting in

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = \frac{1}{2\theta^2}(x^2 - \theta).$$

This form matches with that required of the CRLB theorm and therefore, being that $g(x) = x^2$ is equal to the derived ML estimate $\widehat{\sigma^2}_{MLE}(x) = x^2$ proves that the estimate is efficient.

(e) Use the CRLB theorm to determine whther $\widehat{\sigma}_{MLE}(x)$ is (or is not) and efficient estimator.

Again, following the theory built in the previous section, the derivative of the log-likelihood function is manipulated to determine whether it can be expressed in the form of the CRLB theorm and the derived ML estimate proven to be efficient. Restating the derivative found in part (b)

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = -\frac{1}{\theta} + \frac{x^2}{\theta^3},$$

θ^3 is factored out of the denomonator in a similar manner to that performed in the previous section

$$-\frac{1}{\theta} + \frac{x^2}{\theta^3} = \frac{1}{\theta^3}(x^2 - \theta^2).$$

This form is equivalent to that required by the CRLB theorm however the efficient estimator $g(x) = x^2$ shown here is not equal to the ML estimator derived to be $\widehat{\sigma}_{MLE}(x) = |x|$ proving that this estimator is not efficient.

(f) Given the results of part (c), do the results of parts (d) and (e) suprise you? Explain why or why not.

The relationship between the two estimators in part (c) was shown to differ by a square operation. The results from parts (d) and (e) are not suprising as they highlight an important property of ML estimators: even if an estimator is efficient—as the ML estimator of the variance is—applying a nonlinear transformation to that estimator does not guarantee that the transformed estimator remains efficient. As discussed in Lecture 41 on parameter transformations, an efficient estimator with low variance may yield a transformed estimator that retains many desirable properties, yet efficiency is generally not preserved.

Problem 2 – 25 points

Say that a scalar observation x is related to a source θ as:

$$x = \frac{4\theta}{w}$$

where w is an independent random variable, described by an **exponential *pdf***:

$$p(w) = \begin{cases} \lambda_w e^{-\lambda_w w} & \text{for } w \geq 0 \\ 0 & \text{for } w < 0 \end{cases}$$

The *apriori* knowledge of θ is described by:

$$p(\theta) = \begin{cases} \lambda_\theta e^{-\lambda_\theta \theta} & \text{for } \theta \geq 0 \\ 0 & \text{for } \theta < 0 \end{cases}$$

where $\lambda_\theta > 0$ and $\lambda_w > 0$.

1. Determine the **MLE estimate** $\hat{\theta}_{mle}(x)$.

2. Determine the **MAP estimate** $\hat{\theta}_{map}(x)$.

3. **Evaluate** this MAP estimate for the case where:

$$\lim_{\lambda_\theta \to 0} \hat{\theta}_{map}$$

Explain this result.

4. Now **evaluate** this MAP estimate **instead** for the case where:

$$\lim_{\lambda_w \to 0} \hat{\theta}_{map}$$

Explain this result.

Hint1: Remember EECS 861 !

Hint2: Recall that the **expected value** of a random variable with an exponential distribution is λ^{-1}, and its **variance** is λ^{-2} (these facts will be helpful for your **explanations**!).

(a) Determine the ML estimate $\widehat{\theta}_{MLE}(x)$.

As shown in Lecture 33, the maximum likelihood estimator (MLE) seeks an estimate of θ which maximizes the a derived likelihood function. Given the scalar observation x and its relation to source θ as well as random variable w—denoted as $f(\theta, w)$—the first task in determining the MLE is to determine the distribution of x. This involves a transformation from the given distribution of w described by the property of probability preservation

$$p_X(x)dx = p_W(w)dx$$

which ensures that the probability mass assigned to an interval in w is preserved when mapped to a corresponding interval in x. Being that the transformation $x = f(\theta, w)$ is one-to-one and monotonic, the inverse mapping exists as $f^{-1}(x)$ and therefore the above can be rewritten to

$$p_X(x) = p_W(f^{-1}(x))\left|\frac{df^{-1}(x)}{dx}\right|$$

commonly referred to as the change of variables formula for continuous PDFs where the term $\left|\frac{dw}{dx}\right|$ compensates for stretching or compression of probability density under the mapping. Utilizing the above equation along with the inverse

$$x = f(x, \theta) = \frac{4\theta}{w} \longrightarrow w = f^{-1}(x) = \frac{4\theta}{x},$$

the inverses derivative

$$\frac{dw}{dx} = -\frac{4\theta}{x^2},$$

and the known exponential distribution of w, the PDF of measurement x with conditional dependence on estimation parameter θ is described as

$$p(x|\theta) = \lambda_w \exp\left[-\lambda_w \frac{4\theta}{x}\right]\frac{4\theta}{x^2}.$$

Now, similar to Problem 1, the ML estimate is found by determining and classifying the extreme points of this function. To begin, the natural logarithm is applied as it does effect the location of the stationary points due to its monotonically increasing property and greatly simplifies the distribution expression.

$$\ln(p(x|\theta)) = \ln(\lambda_w) + \ln\left(\frac{4\theta}{x^2}\right) - \frac{4\lambda_w\theta}{x} = \ln(\lambda_w) + \ln(4) + \ln(\theta) - \ln(x^2) - \frac{4\lambda_w\theta}{x}$$

The derivative with respect to estimation varaible θ is

$$\frac{\partial \ln(p(x|\theta))}{\partial\theta} = \frac{1}{\theta} - \frac{4\lambda_w}{x}$$

which when set equal to zero and solving for θ results in the stationary point

$$\frac{1}{\theta} - \frac{4\lambda_w}{x} = 0 \quad \longrightarrow \quad \theta = \frac{x}{4\lambda_w}.$$

Utilizing the second derivative test to determing if this point is a global maximum yeilds

$$\frac{\partial^2 \ln(p(x|\theta))}{\partial^2\theta} = -\frac{1}{\theta^2} < 0 \ \forall \ x$$

proving that $\theta = \frac{x}{4\lambda_w}$ is the MLE.

$$\widehat{\theta}_{MLE}(x) = \frac{x}{4\lambda_w}$$

(b) Determine the MAP estimate $\widehat{\theta}_{MAP}(x)$**.**

Utilizing the likelihood function derived in the previous section, the maximum a posteriori estimate is that which maximizes the a posteriori density function defined as the product between the likelihood and a piori densities as described by Bayes theorm.

$$p(\theta, x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Here, the demononator density of measurement x is omitted as it deterministically scales the function and is not required as only the source θ is to be estimated, the not the value of the maximum this parameter produces. Given the a priori information of source θ being exponentially distributed, the a posteriori density function is written as

$$p(\theta|x) = p(x|\theta)p(\theta) = \left(\lambda_w \exp\left[-\lambda_w \frac{4\theta}{x}\right] \frac{4\theta}{x^2}\right)(\lambda_\theta \exp[-\lambda_\theta \theta])$$

which is rearranged to

$$p(\theta|x) = \frac{4\lambda_w \lambda_\theta \theta}{x^2} \exp\left[-\frac{4\lambda_w \theta}{x} - \lambda_\theta \theta\right].$$

Determining the value of θ that maximizes this function is preformed similar to that in the previous section where first the derivative is taken and its extreme points classified. Applying the logarithm operator to the above density function and taking the derivative with respect to θ yeilds

$$\ln(p(\theta|x)) = \ln(4\lambda_w\lambda_\theta) + \ln(\theta) - \ln(x^2) - \frac{4\lambda_w\theta}{x} - \lambda_\theta\theta$$

$$\frac{\partial \ln(p(\theta|x))}{\partial \theta} = \frac{1}{\theta} - \frac{4\lambda_w}{x} - \lambda_\theta.$$

This result is set equal to zero and solved for estimation parameter θ below.

$$\frac{1}{\theta} - \frac{4\lambda_w}{x} - \lambda_\theta = 0 \quad \longrightarrow \quad \theta = \frac{1}{\lambda_\theta + \frac{4\lambda_w}{x}}$$

Similar to that in part (a), the second derivative test is performed to validate that this stationary point is a global maximum

$$\frac{\partial^2 \ln(p(\theta|x))}{\partial^2 \theta} = -\frac{1}{\theta^2} < 0 \;\; \forall \;\; x$$

which proves that the a posteriori density function is concave and the drived value for θ is the MAP estimate.

$$\widehat{\theta}_{MAP} = \frac{1}{\lambda_\theta + \frac{4\lambda_w}{x}}$$

(c) Evaluate this MAP estimate for the case where $\lim\limits_{\lambda_\theta \to 0} \widehat{\theta}_{MAP}$***.***

In the case in which the a priori parameter λ_θ decreases towards zero, the resulting mean—$\mathbb{E}\{\theta\} = \lambda_\theta^{-1}$—of the a priori distribution increases towards ∞ resulting in the a priori distribution be approximately flat and providing no additional information about the estimation parameter θ. For the estimator found in part (b), plugging in $\lambda_\theta \to 0$ yeilds

$$\widehat{\theta}_{MAP} = \frac{1}{0 + \frac{4\lambda_w}{x}} = \frac{x}{4\lambda_w} = \widehat{\theta}_{MLE}$$

where without a priori information the estimate becomes the MLE. This limit matches exactly with what is expected in Bayesian estimation theory in which the maximum likelihood estimate is a result of the a priori contributing no useable information.

(d) Now evaluate this MAP estimate instead for the case where $\lim\limits_{\lambda_w \to 0} \widehat{\theta}_{MAP}$***.***

For the case in which the random variable w's distribution parameter λ_w decreases towards zero, again the resulting mean of this parameter converges towards ∞ which, being that measurement x is inversely proportional to this parameter, will result in a measurement that is extremely small. Plugging $\lambda_w \to 0$ into the estimate derived in part (b) yeilds

$$\widehat{\theta}_{MAP} = \frac{1}{\lambda_\theta + 0} = \frac{1}{\lambda_\theta}$$

which is interpreted as with no additional information provided by the measurement, all useable information is a result of the a priori. Again this matches strongly to Bayesian estimation theory where information is contributed both through the likelihood function—dependent on the measurement—and the a priori information, when one of these becomes uninformative the resulting estimate will fault to the other.

Problem 3 - 25 points

Say the elements x_n of an **N-dimensional** vector $\mathbf{x}$ are **independent** and identically distributed random variables, each with an Exponential *pdf* :

$$p(x_n|\lambda) = \lambda\, exp(-\lambda x_n) \quad \text{for } x_n > 0$$

Therefore,

$$\begin{aligned} p(\mathbf{x}|\lambda) &= \prod_{n=1}^{N} p(x_n|\lambda) \\ &= \prod_{n=1}^{N} \lambda\, exp(-\lambda x_n) \quad \text{for } x_n > 0 \end{aligned}$$

Determine the **MLE estimate** of λ (i.e., determine $\hat{\lambda}(\mathbf{x})$).

Hint: Recall that $e^x\, e^y = e^{x+y}$.

***(a) Determine the ML estimate of* λ (i.e. determine $\widehat{\lambda}_{MLE}(\mathbf{x})$).**

Given the joint likelihood function of distribution parameter λ producing measurement vector $\mathbf{x}$ as

$$p(\mathbf{x}|\lambda) = \prod_{n=1}^{N} \lambda \exp[-\lambda x_n],$$

the task of determining the maximum likelihood (ML) estimate is that of finding the value of λ which maximizes this distribution. Utilizing the exponential power rule of $e^x e^y = e^{x+y}$, the above distribution can be alternately expressed as

$$p(\mathbf{x}|\lambda) = \lambda^N \exp\left[-\lambda \sum_{n=1}^{N} x_n\right].$$

The next step is to take the gradient of the multivariate distribution with respect to the estimation parameter λ which is first preceeded by applying the natural logarithm operator to algebraically simplify subsequent operations.

$$\ln(p(\mathbf{x}|\lambda) = N\ln(\lambda) - \lambda \sum_{n=1}^{N} x_n$$

$$\frac{\partial \ln(p(\mathbf{x}|\lambda))}{\partial \lambda} = \frac{N}{\lambda} - \sum_{n=1}^{N} x_n$$

Setting the result above equal to zero and solving for λ yeilds

$$\frac{N}{\lambda} - \sum_{n=1}^{N} x_n = 0 \quad \longrightarrow \quad \lambda = N\left(\sum_{n=1}^{N} x_n\right)^{-1}$$

To classify this extrema the second derivative test is used

$$\frac{\partial^2 \ln(p(\mathbf{x}|\lambda))}{\partial^2 \lambda} = -\frac{N}{\lambda^2} < 0 \;\; \forall \;\; \mathbf{x},$$

which, as shown above, is less than zero for all measurement vectors $\mathbf{x}$. This proves that the function is concave, the stationary point found is a global maximum, and being that this value of λ maximizes the likelihood function, it is the ML estimate.

$$\widehat{\lambda}_{MLE}(\mathbf{x}) = N\left(\sum_{n=1}^{N} x_n\right)^{-1}$$