(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.

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

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.

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

(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

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

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 $\lambda_w \rightarrow 0$ 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.