(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.

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

(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

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

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

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

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

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

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