Algorithms & DATA - COMS20017

Tutorial Questions & Solutions

(Q1) In the context of Minimum Variance Unbiased Estimation (MVUE), define the likelihood function for a random signal x. State the likelihood function for a single random sample $x[0] = A + w[0]$, where A is the DC level and $w \sim \mathcal{N}(0, \sigma^2)$

Solution: The likelihood function is essentially the same as the PDF, but seen as a function of the unknown parameter rather than of the dataset x, which is fixed.

For the given random variable, the log-likelihood function is

$$\ln p(x[0]; A) = -\ln \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\left(x[0] - A\right)^2$$

(Q2) Define the curvature of the log-likelihood function. What information does the curvature provide? Explain why it is convenient to use the log-likelihood function.

Solution: The "sharpness" of the likelihood function determines how accurately one can estimate the unknown parameter.

Mathematically, the sharpness is measured by the negative of the second derivative of the log-likelihood function.

Because the second derivative generally also depends on $x[0]$, the measure of curvature that is normally employed is

$$-E\left[\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}\right]$$

which measures the average curvature of the log-likelihood function.

Using the log-likelihood has advantages in terms of mathematical tractability, since by applying the logarithm, products are converted into sums and exponentials are avoided.

(Q3) Define and discuss the Cramer-Rao Lower Bound (CRLB) for scalar parameters.

Solution: Assuming that the PDF satisfies the regularity condition

$$E\left[\frac{\partial \ln p(x; A)}{\partial A}\right] = 0 \quad \forall A,$$

then the variance of any unbiased estimator $\hat{A}$ must satisfy

$$Var(\hat{A}) \geq \frac{1}{-E\left[\dfrac{\partial^2 \ln p(x; A)}{\partial A^2}\right]}$$

An unbiased estimator can be found that attains the bound for all A iff

$$\frac{\partial \ln p(x;A)}{\partial A} = I(A)(g(x) - A)$$

for some functions g and I. That estimator, which is the MVU estimator, is $\hat{A} = g(x)$ and the minimum variance is $\frac{1}{I(A)}$.

(Q4) In the context of estimating a DC level in white Gaussian noise, consider N observations

$$x[n] = A + w[n], \quad n = 0, 1, 2, \dots, N-1,$$

where $w[n] \sim \mathcal{N}(0, \sigma^2)$. Given that

$$p(x;A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right\}$$

determine the CRLB for A.

Solution: Take the first derivative of the (given) log-likelihood function to yield

$$\begin{aligned} \frac{\partial \ln p(x;A)}{\partial A} &= \frac{\partial}{\partial A}\left[-\ln(2\pi\sigma^2)^{N/2} - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) = \frac{N}{\sigma^2} (\bar{x} - A) \end{aligned}$$

where $\bar{x}$ is the sample mean.

Differentiate once again to have

$$\frac{\partial^2 \ln p(x;A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

Finally, the CRLB is given as $Var(\hat{A}) \geq \frac{\sigma^2}{N}$.

(Q5) Let X denote a Poisson random variable with probability density function

$$f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{for} \quad x = 0, 1, \dots.$$

Assuming that the rate parameter λ is exponentially distributed with

$$f(\lambda) = \frac{1}{\lambda_0} e^{-\frac{\lambda}{\lambda_0}}$$

and the joint density of x and λ is

$$f(x, \lambda) = f(x|\lambda) f(\lambda)$$

determine the maximum a posteriori estimate of λ and comment on the values of λ when λ_0 is much smaller than 1.

Solution: The MAP estimator is the mode of the posterior pdf:

$$\hat{\lambda}(x) = \arg\max_{\lambda} f(\lambda|x)$$

Using Bayes theorem together with the above equation, one gets

$$\begin{aligned} \hat{\lambda} &= \arg\max_{\lambda} f(x|\lambda) f(\lambda) \\ &= \arg\max_{\lambda} [\ln f(x|\lambda) + \ln f(\lambda)] \end{aligned}$$

Plugging the given likelihood and prior pdf's in the last equation above

$$\hat{\lambda} = \arg\max_{\lambda} \left[x \log \lambda - \log(\lambda_0 x!) - \lambda \left(1 + \frac{1}{\lambda_0} \right) \right]$$

The MAP estimate is obtained by taking the derivative of the above expression and setting it equal to 0:

$$\frac{x}{\lambda} - \left(1 + \frac{1}{\lambda_0} \right) = 0 \quad \rightarrow \quad \hat{\lambda}_{MAP} = \frac{\lambda_0}{\lambda_0 + 1} \cdot x.$$

In cases when $\lambda_0 \ll 1$, the denominator of the MLE will be $(\lambda_0 + 1)|_{\lambda_0 \ll 1} = 1$. Then,

$$\hat{\lambda}_{MLE} = \lambda_0 \cdot x.$$

(Q6) Let x denote a Rayleigh distributed random variable with probability density function given by

$$f(x|\theta) = \frac{x}{\theta^2} \exp\left\{ -\frac{x^2}{2\theta^2} \right\}$$

Determine the maximum likelihood estimate of θ.

Solution: The likelihood function can be written as

$$\begin{aligned} p(x|\theta) &= \prod_{n=1}^{N} f(x_i|\theta) \\ &= \frac{\prod_{n=1}^{N} x_i}{\theta^{2N}} \exp\left[-\frac{1}{2\theta^2} \sum_{n=1}^{N} x_i^2 \right] \end{aligned}$$

Log-likelihood function is

$$L(\theta|x) = \sum_{n=1}^{N} \log x_i - 2N \log \theta - \frac{1}{2\theta^2} \sum_{n=1}^{N} x_i^2$$

Taking the derivative yields

$$\frac{\partial L(\theta|x)}{\partial \theta} = -\frac{2N}{\theta} + \frac{2}{2\theta^3} \sum_{n=1}^{N} x_i^2$$

The ML estimate of θ is obtained by setting to zero the derivative of L w.r.t. θ

$$\frac{\partial L(\theta|x)}{\partial \theta} = 0 \quad \rightarrow \quad \theta^2 = \frac{1}{2N} \sum_{n=1}^{N} x_i^2$$

which leads to the ML estimate of θ

$$\hat{\theta}_{MLE} = \sqrt{\frac{1}{2N} \sum_{n=1}^{N} x_i^2}$$

(Q7) Consider the problem of estimating a DC level A in white Gaussian noise, w, where the noisy data are given by

$$x[n] = A + w[n], \quad n = 0, \ldots, N-1, \quad w[n] \sim \mathcal{N}(0, \sigma_w^2)$$

(a) Estimate the value of A using the maximum likelihood estimation (MLE) procedure. Discuss briefly the optimality properties of the MLE.

(b) Estimate the value of A using the method of least squares (LS). Discuss briefly the properties of LS estimation. State at least one problem associated with this approach.

(c) Determine the Cramer Rao lower bound (CRLB) of the unknown parameter A. Compare this solution with those from (a) and (b)

<u>Solution:</u>

(a) PDF of x is

$$p\left(x, A\right) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left(x\left[n\right] - A\right)^2\right]$$

Log-likelihood function:

$$\ln p\left(x, A\right) \sim -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(x\left[n\right] - A\right)^2$$

Derivative of log-likelihood function:

$$\frac{\partial \ln p(x, A)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)$$

Equating to zero and solving yields the MLE:

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

If an efficient estimator exists, then the MLE procedure will produce it.

When an efficient estimator does not exist, the MLE yields an asymptotically efficient estimator, i.e. one that for large datasets

- is unbiased
- achieves the CRLB
- has a Gaussian PDF, $\hat{\theta} \sim N\left(\theta, I^{-1}(\theta)\right)$

(b)

$$J(A) = \sum_{n=0}^{N-1} (x[n] - A)^2$$

$$\frac{\partial J(A)}{\partial A} = 2 \sum_{n=0}^{N-1} (x[n] - A)$$

Set the result to zero to yield the LSE

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

In LS estimation, no assumptions are necessary regarding the pdf of the noise.

The estimate is again represented my the familiar sample mean, however it cannot be claimed to be optimal in the MVU sense, except if $x[n] = A + w[n], \quad \text{with} \quad w[n] \sim N(0, \sigma_w^2)$; all that can be said is that it minimises the sum squared errors.

This estimation can only be applied to zero-mean noise.

(c)

$$\begin{aligned} p(x, A) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (x[n] - A)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right] \end{aligned}$$

Take the first derivative

$$\begin{aligned}\frac{\partial \ln p\,(x, A)}{\partial A} &= \frac{\partial}{\partial A}\left[-\ln\left(2\pi\sigma^2\right)^{N/2} - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x\,[n]-A\right)^2\right] \\ &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x\,[n]-A\right) = \frac{N}{\sigma^2}\left(\bar{x}-A\right)\end{aligned}$$

where $\bar{x}$ is the sample mean. Take the second derivative

$$\frac{\partial^2 \ln p\,(x, A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

Therefore, $Var\left(\hat{A}\right) \geq \dfrac{\sigma^2}{N}$ is the CRLB. This implies that the sample mean estimator, which is the estimator obtained in both (a) and (b), attains the bound and consequently is the MVUE in white Gaussian noise.

(Q8) (a) Derive the Cramer-Rao lower bound (CRLB) for the estimation of a DC level in white Gaussian noise ($w[n] \sim \mathcal{N}(0, \sigma^2)$), given by

$$x[n] = A + w[n], \quad n = 0, 1, \ldots, N-1$$

For which the probability density function is given by

$$p(x; A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right]$$

(b) Derive the maximum likelihood estimator (MLE) for the problem from Q8 (a).

(c) Is the MLE from Q8 (b) unbiased? Is it efficient? Does such an MLE attain the CRLB?

Solution:

(a) The CRLB for A can be obtained as follows

$$\begin{aligned}p(x; A) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right] \\ p(x; A) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]\end{aligned}$$

Taking the first derivative

$$\begin{aligned}\frac{\partial \ln p(x;A)}{\partial A} &= \frac{\partial}{\partial A}\left[-\ln(2\pi\sigma^2)^{N/2} - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) \\ &= \frac{N}{\sigma^2}(\bar{x}-A)\end{aligned}$$

where $\bar{x}$ is the sample mean. Taking the second derivative, we have

$$\frac{\partial^2 \ln p(x;A)}{\partial A^2} = \frac{N}{\sigma^2}$$

Hence, $Var(\hat{A}) = \frac{\sigma^2}{N}$ is the CRLB.

(b) The pdf has absolutely the same form as from point (b) above, therefore the derivative of the log-likelihood function will be

$$\frac{\partial \ln p(x;A)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)$$

Setting the result to zero yields the MLE

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1}x[n] = \bar{x}.$$

(c) In this example it is obvious that the MLE represents the MVUE, which yields the CRLB and hence it is efficient.

If an efficient estimator does exist, then the MLE procedure will produce it.

An essential property of the MLE is that it yields an asymptotically efficient estimator, i.e. one that for sufficiently large sample sizes is unbiased and attains the CRLB

(Q9) Derive the general expression for the maximum a posteriori (MAP) estimator of θ, given the observed variable x.

Solution: The Bayesian estimator $\hat{\theta}$ minimizes the conditional risk, which is the loss (cost function) averaged over the conditional distribution of θ, given the observation x:

$$\hat{\theta}(x) = \arg\min_{\theta}\int C[\theta,\hat{\theta}(x)]p(\theta|x)d\theta$$

By selecting the uniform cost function, the MAP is obtained as

$$\begin{aligned}\hat{\theta}(x) &= \arg\min_{\theta} \int_{|\theta-\hat{\theta}|\geq\delta} p(\theta|x)d\theta \\ &= \arg\min_{\theta} \left[1 - \int_{|\theta-\hat{\theta}|\geq\delta} p(\theta|x)d\theta\right]\end{aligned}$$

In order to minimize the expected cost, when $\delta \to 0$ one should select

$$\hat{\theta}(x) = \arg\max_{\theta} p(\theta|x)$$

that is, the mode of the posterior pdf. Using Bayes theorem together with the above equation, one gets

$$\hat{\theta}(x) = \arg\max_{\theta} p(x|\theta)p(\theta)$$

Or

$$\hat{\theta}(x) = \arg\max_{\theta} [\ln p(x|\theta) + \ln p(\theta)]$$

(Q10) (a) Let $x[n]$ represent N measurements of a constant amplitude signal. The measurement is corrupted by white Gaussian noise with zero mean and variance σ^2. If

$$x[n] = A + w[n] \quad \text{for} \quad n = 0, 1, \ldots, N-1,$$

show that the Maximum Likelihood estimate for A is given by

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n].$$

(b) In the same model as above Q10 (a), consider now that A is a zero-mean Gaussian random variable i.e.

$$p(A) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{A^2}{2\sigma^2}\right)$$

Assuming that the likelihood function is Gaussian as well,

$$p(x|A) = \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left(-\frac{(x-A)^2}{2\sigma_n^2}\right)$$

Derive the MAP estimate of A.

(c) For the same model as in (a) and (b) above, derive now the MMSE estimate of A.

Solution:

(a) The likelihood function is

$$p\left(x|A\right) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{1}{2\sigma_n^2}\left(x\left[n\right]-A\right)^2\right]$$

The log likelihood function gives

$$L(x, A) = -\frac{N}{2}\ln(2\pi\sigma_n^2) - \frac{1}{2\sigma_n^2}\sum_{n=0}^{N-1}\left(x\left[n\right]-A\right)^2$$

The ML estimate of A is obtained by setting to zero the derivative of $L(\cdot)$ w.r.t. A

$$\frac{\partial L(x, A)}{\partial A} = \frac{1}{2\sigma_n^2}\sum_{n=0}^{N-1}\left(x\left[n\right]-A\right)^2 = 0$$

which then gives the ML estimate of A as

$$\hat{A}_{MLE} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

(b) The MAP estimator is the mode of the posterior pdf:

$$\hat{A}(x) = \arg\max_A p(A|x)$$

Using Bayes theorem together with the above equation, one gets

$$\begin{aligned}\hat{A} &= \arg\max_A p(x|A)p(A)\\ &= \arg\max_A \left[\ln p(x|A) + \ln p(A)\right]\end{aligned}$$

Plugging the given likelihood and prior PDF's in the last equation above

$$\hat{A} = \arg\max_A \left[-\frac{(x-A)^2}{2\sigma_n^2} - \frac{A^2}{2\sigma^2}\right]$$

Differentiating with respect to A yields

$$\frac{\partial}{\partial A}\left[-\frac{(x-A)^2}{2\sigma_n^2} - \frac{A^2}{2\sigma^2}\right] = -\frac{x-A}{\sigma_n^2} - \frac{A}{\sigma^2}$$

and setting it equal to 0 yields

$$\hat{A} = \frac{\sigma^2}{\sigma^2+\sigma_n^2}\cdot x$$

(c) The MMSE is the mean of the posterior pdf:

$$\hat{A}_{MMSE} = \mathbb{E}[A|x] = \mu_{A|x}$$

Using Bayes theorem, the posterior distribution is

$$P(A|x) \propto p(x|A)p(A)$$

and substituting the given likelihood and prior:

$$P(A|x) \propto exp\left(-\frac{(x-A)^2}{2\sigma_n^2} - \frac{A^2}{2\sigma^2}\right)$$

Working on the exponent:

$$\begin{aligned}
&-\frac{(x-A)^2}{2\sigma_n^2} - \frac{A^2}{2\sigma^2} = \\
&= -\frac{x^2 - 2Ax + A^2}{2\sigma_n^2} - \frac{A^2}{2\sigma^2} = \\
&= -\frac{A^2}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right) + \frac{Ax}{\sigma_n^2} - \frac{x^2}{2\sigma_n^2}
\end{aligned}$$

The obtained expression is a quadratic in A, which is indicative of a Gaussian distribution. We will now ignore the term $-\frac{x^2}{2\sigma_n^2}$, which can be made part of the normalization constant of the Gaussian, and work only on the first two terms:

$$\begin{aligned}
&-\frac{A^2}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right) + \frac{Ax}{\sigma_n^2} = \\
&= -\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)\left(A^2 - \frac{2Ax}{\sigma_n^2\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)}\right) = \\
&= -\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)\left(A - \frac{x}{\sigma_n^2\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)}\right)^2 - \left(\frac{x}{\sigma_n^2\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)}\right)
\end{aligned}$$

where the last term in parentheses, which was added and then subtracted to complete the square, can again be dropped and made part of the normalizing constant of the Gaussian distribution. Hence, we have for the posterior distribution:

$$P(A|x) \propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)\left(A - \frac{x}{\sigma_n^2\left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma^2}\right)}\right)^2\right)$$

Finally, identifying the terms with the expression of a generic Gaussian distribution, we obtain the posterior variance as:

$$\sigma_{A|x}^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_n^2}\right)^{-1} = \frac{\sigma^2\sigma_n^2}{\sigma^2 + \sigma_n^2}$$

and the posterior mean as:

$$\mu_{A|x} = \frac{\sigma^2}{\sigma^2 + \sigma_n^2} x$$

Observe that the result for the Bayesian MMSE is identical to that obtained in point (b) for the MAP. This is expected given the symetry of all the distributions involved, i.e. in this case the mode and and the mean of the posterior distribution are the same.

(Q11) (a) Assume that the likelihood function is Gaussian, i.e.

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

and that the prior pdf is Cauchy:

$$p(\theta) = \frac{\gamma}{\pi(\gamma^2 + \theta^2)}$$

Show that the MAP estimate of θ can be obtained in closed-form.

(b) Now assume that the likelihood function is still Gaussian, as in Q11 (a), but this time the prior pdf is Laplace:

$$p(\theta) = \frac{1}{\sqrt{2}\sigma} \exp\left\{ -\frac{\sqrt{2}|\theta|}{\gamma} \right\}$$

Show that the MAP estimate of θ can be obtained in closed-form.

<u>**Solution:**</u> The MAP estimator is the mode of the posterior pdf:

$$\hat{\theta}(x) = \arg\max_{\theta} p(\theta|x)$$

Using Bayes theorem together with the above equation, one gets

$$\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} p(x|\theta)p(\theta) \\
&= \arg\max_{\theta} \left[\ln p(x|\theta) + \ln p(\theta)\right]
\end{aligned}$$

(a) Plugging the given likelihood function and prior pdf in the last equation above

$$\hat{\theta}(x) = \arg\max_{\theta} \left[-\frac{(x-\theta)^2}{2\sigma^2} + \ln\frac{\gamma}{\pi(\theta^2+\gamma^2)}\right]$$

Differentiating with respect to θ

$$\frac{\partial}{\partial\theta}\left[-\frac{(x-\theta)^2}{2\sigma^2} + \ln\frac{\gamma}{\pi(\theta^2+\gamma^2)}\right] = \frac{(x-\theta)}{\sigma^2} - \frac{2\theta}{\theta^2+\gamma^2}$$

and setting it equal to 0 yields

$$\frac{(x-\theta)}{\sigma^2} - \frac{2\theta}{\theta^2+\gamma^2} = 0$$
$$\theta^3 - x\theta^2 + (\gamma^2 + 2\sigma^2)\theta - \gamma^2 x = 0$$

which is a third order equation in θ and thus $\hat{\theta}$ can be obtained in closed form.

(b) Plugging the given likelihood function and prior pdf in the expression for the MAP estimator, we get:

$$\hat{\theta}(x) = \arg\max_{\theta}\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{\sqrt{2}|\theta|}{\gamma}\right]$$

Differentiating with respect to θ

$$\frac{\partial}{\partial\theta}\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{\sqrt{2}|\theta|}{\gamma}\right] = \frac{(x-\theta)}{\sigma^2} - \frac{\sqrt{2}sign(\theta)}{\gamma}$$

and setting it equal to 0 yields

$$\hat{\theta} = x - \frac{\sqrt{2}\sigma^2}{\gamma}sign(\theta)$$
$$= sign(x)\left(|x| - \frac{\sqrt{2}\sigma^2}{\gamma}\right)_+$$

where $(g)_+$ is defined as $(g)_+ = \begin{cases} 0, & \text{if } g < 0 \\ g, & \text{otherwise} \end{cases}$.