E1 244: Detection and Estimation Theory
Spring 2011 – Test 1 Solutions

1. Conditioned on θ, a random variable Y has Gaussian distribution with unknown mean θ and known variance σ^2. It is known that prior distribution of θ itself is Gaussian with known mean θ_0 and known variance σ_0^2.

   (a) Find the MMSE estimator of θ when a single sample of Y is observed.
   Solution:

$$\begin{aligned} p_\theta(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\theta)^2}{2\sigma^2}\right] \\ w(\theta) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right] \\ w(\theta|y) &= \frac{\frac{1}{2\pi\sigma_0\sigma}\exp\left[-\frac{(\theta-\mu_0)^2}{2\sigma_0^2} - \frac{(y-\theta)^2}{2\sigma^2}\right]}{\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_0\sigma}\exp\left[-\frac{(\theta-\mu_0)^2}{2\sigma_0^2} - \frac{(y-\theta)^2}{2\sigma^2}\right] d\theta} \end{aligned}$$

   After completing the square inside exponential term and some manipulation,

$$w(\theta|y) = \frac{1}{\sqrt{2\pi\frac{\sigma_0^2\sigma^2}{\sigma_0^2+\sigma^2}}} \exp\left[-\frac{\left(\theta - \frac{y\sigma_0^2+\mu_0\sigma^2}{\sigma_0^2+\sigma^2}\right)^2}{2\frac{\sigma_0^2\sigma^2}{\sigma_0^2+\sigma^2}}\right]$$

   Hence,

$$\hat{\theta}_{\text{MMSE}}(y) = \mathbb{E}_\theta w(\theta|y) = \frac{y\sigma_0^2 + \mu_0\sigma^2}{\sigma_0^2+\sigma^2}$$

   (b) What happens when $\sigma_0^2 \to 0$?
   Solution: As $\sigma_0^2 \to 0$, $\hat{\theta}_{\text{MMSE}}(y) \to \mu_0$. This is intuitive since as $\sigma_0^2 \to 0$, θ takes value μ_0 with probability 1.

   (c) What happens when $\sigma_0^2 \to \infty$? What kind of estimator is this?
   Solution: As $\sigma_0^2 \to \infty$, $\hat{\theta}_{\text{MMSE}}(y) \to y$. This is nothing but the Maximum Likelihood Estimate, since $\sigma_0^2 \to \infty$ can be thought of as having no prior knowledge about θ.

2. **Random Access.** In a given slot (say, an hour), the number K of students accessing IEEE Xplore is Poisson distributed with unknown mean λ. Each time a student access

IEEE Xplore, there is a probability p, also unknown, that the web-site is available, and with probability $1-p$, the student will get the message *"All available online seats are currently occupied"*. As system administrator, you want to estimate the parameters λ and p. For some slot, you record $K = k$ students accessing IEEE Xplore, and their outcomes $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_k]$, where $Y_i = 1$ if the access was successful (i.e., the website was available) and $Y_i = 0$ otherwise.

(a) Find the sufficient statistics (of the smallest dimension) for the parameter $\theta = (p, \lambda)$.
Solution: The pdf of the observation given θ is

$$p_\theta(\mathbf{y}, k) = e^{-\lambda}\frac{\lambda^k}{k!}\left(\frac{p}{1-p}\right)^{\sum_{i=1}^k y_i}(1-p)^k$$

By the NF factorization theorem, $(k, \sum_{i=1}^k y_i)$ is sufficient.

(b) Assuming p is known, find the ML estimator $\hat{\lambda}_{ML}$ of λ. Is the estimator unbiased? What is the MSE of the ML estimator?
Solution: When p is known, the log likelihood function is

$$l_\lambda(\mathbf{y}, k) = -\lambda + k\log\lambda + g$$

where g contains terms that are independent of λ. The ML estimate of λ is then given by the solution to

$$-1 + \frac{k}{\lambda} = 0,$$

which yields

$$\hat{\lambda}_{ML} = k.$$

Since K is Poission distributed with mean λ, the ML estimator is unbiased. Moreover, its MSE (variance) is λ as well.

(c) Suppose neither λ nor p is known. Does a UMVUE estimator for p exist? If so, find it. If not, show why not.
Solution: The log-likelihood function is given by

$$l_\theta(\mathbf{y}, k) = h(k, \lambda) + \left(\sum_{i=1}^k y_i\right)\log\left(\frac{p}{1-p}\right) + k\log(1-p)$$

Writing the pdf as an exponential form,

$$p_\theta(\mathbf{y}, k) = \exp\left\{k\log\left((1-p)\lambda\right) + \left(\sum_{i=1}^k y_i\right)\log\frac{p}{1-p} - \lambda - \log k!\right\},$$

from which $\left(k, \sum_{i=1}^k y_i\right)$ is a complete sufficient statistic. Now consider ML estimation of p. From the log-likelihood function,

$$\hat{p}_{ML} = \frac{1}{K}\sum_{i=1}^K Y_i$$

The ML estimate is unbiased, because

$$\mathbb{E}\hat{p}_{ML} = \mathbb{E}\left(\mathbb{E}\left(\frac{1}{K}\sum Y_i|K\right)\right) = p$$

Thus, the ML estimate is unbiased estimate of p, and a function of the complete sufficient statistic. Hence, it is also the UMVUE of p.

3. **Mixed Cost Function.** Consider the following model:

$$\mathbf{y} = A\mathbf{w} + \mathbf{n}, \tag{1}$$

where

- $\mathbf{y} \in \mathbb{R}^{n\times 1}$ is the observation vector
- $A \in \mathbb{R}^{n\times n}$ is a known data matrix whose entries are $[\mathbf{h}_1^T, \mathbf{h}_2^T, \dots, \mathbf{h}_n^T]^T$, with $h_i \in \mathbb{R}^{n\times 1}$
- $\mathbf{w} \in \mathbb{R}^{n\times 1}$ is an unknown weight vector
- $\mathbf{n} \in \mathbb{R}^{n\times 1}$ is a noise vector, also unknown

(a) Find a Least Squares (LS) estimate of $\mathbf{w}$ given the observation. Make any reasonable assumption on the matrix A that you may need.

Solution: The LS problem can be stated as:

$$\min_{\mathbf{w}} (\mathbf{y} - A\mathbf{w})^T(\mathbf{y} - A\mathbf{w}).$$

Writing the above as a quadratic function of $\mathbf{w}$, we have,

$$J(\mathbf{w}) = \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T A\mathbf{w} + \mathbf{w}^T A^T A\mathbf{w}. \tag{2}$$

Taking the gradient and equating to zero results in

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = -2A^T\mathbf{w} + 2A^T A\mathbf{w} = 0. \tag{3}$$

This results in $\hat{\mathbf{w}}_{LS} = (A^TA)^{-1}A^T\mathbf{y}$. Note that the matrix A should have full rank in order for the LS problem to have unique solution.

(b) Find an MMSE estimate of $\mathbf{w}$ if $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 I_{n\times n})$, and the entries of $\mathbf{w}$ are i.i.d. $\mathcal{N}(0, \sigma_w^2)$ and independent of $\mathbf{n}$.

Solution: Since $\mathbf{w}$ and $\mathbf{y}$ are jointly Gaussian, defining $\mathbf{z} = (\mathbf{w}^T, \mathbf{y}^T)^T$, we have that

$$\mathbf{Z} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_w^2 I & \sigma_w^2 A^T \\ \sigma_w^2 A & \sigma_w^2 AA^T + \sigma_n^2 I \end{bmatrix}\right).$$

Hence, the probability density function of $\mathbf{w}$ conditioned on $\mathbf{y}$ is Gaussian with mean

$$\mu_{\mathbf{w}|\mathbf{y}} = \sigma_w^2 A^T \left(\sigma_w^2 AA^T + \sigma_n^2 I\right)^{-1}\mathbf{y},$$

Finally, since the conditional mean is the MMSE estimate, $\hat{\mathbf{w}}_{MMSE} = \mu_{\mathbf{w}|\mathbf{y}}$.

(c) Let

$$\rho(r_i) \triangleq \begin{cases} r_i^2/2, & |r_i| \le b \\ b\,|r_i| - b^2/2, & |r_i| > b \end{cases} \tag{4}$$

for some $b \in \mathbb{R}$.

i. Plot $\rho(r_i)$. Is it a convex function of r_i?
Solution: The plot is simple to draw. It is clear that the function is linear for $|r_i| > b$, and hence there exist two points r_{1i} and r_{2i} such that the line joining the two lies below the curve. Thus, it is a non convex function in r_i.

ii. Find the derivative of $\rho(r_i)$.
Solution: The derivative can be easily shown to be

$$\frac{\partial\rho(r_i)}{\partial r_i} = \begin{cases} r_i, & |r_i| \le b \\ b\ \mathrm{sgn}(r_i), & |r_i| > b \end{cases} \tag{5}$$

iii. Define $r_i \triangleq y_i - \mathbf{h}_i^T\mathbf{w}$, where $\mathbf{y}_i$ is the i-th component of $\mathbf{y}$. Also, let $\mathbb{E}(\mathbf{n}\mathbf{n}^T) = diag[\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2]$, and $\mathbb{E}(\mathbf{n}) = 0$, where $\sigma_1, \ldots \sigma_n$ are known. Assuming that the entries of the noise vector conditioned on the observation are independent, for the model in (1), find an estimate of $\mathbf{w}$ that minimizes the cost function $J(\mathbf{w}) = \sum_{i=1}^n \rho(r_i/\sigma_i)$, where $\rho(r_i)$ is as defined in (4).
Hint: You may need an iterative solution.
Solution: Define $q(r_i/\sigma_i) \triangleq \frac{1}{r_i}\frac{\partial\rho(r_i/\sigma_i)}{\partial r_i}$. Taking the derivative of the cost function and equating it to zero gives:

$$\sum_{i=1}^n \mathbf{h}_i r_i q(r_i/\sigma_i) = \sum_{i=1}^n \mathbf{h}_i(\mathbf{y}_i - \mathbf{h}_i^T\mathbf{w}).q(r_i/\sigma_i) = 0 \tag{6}$$

The above equation can be rewritten as follows:

$$\sum_{i=1}^n \mathbf{h}_i q(r_i/\sigma_i)\mathbf{y}_i = \sum_{i=1}^n \mathbf{h}_i\mathbf{h}_i^T q(r_i/\sigma_i)\mathbf{w}. \tag{7}$$

Writing the above in matrix form as $A^TQ\mathbf{y} = A^TQA\mathbf{w}$, which results in $\hat{w} = (A^TQA)^{-1}A^TQ\mathbf{y}$, where the matrix Q is defined as follows:

$$Q = \mathrm{diag}(q(r_1/\sigma_1), q(r_2/\sigma_2), \ldots, q(r_n/\sigma_n)). \tag{8}$$

We start with an initial guess for $\mathbf{w}$ and iteratively find the optimum $\mathbf{w}$.