EE 631: Estimation and Detection
Part 5

Dr. Mehrdad Soumekh

Parameter Estimation (contd.)

In the last lecture, it was shown that the efficiency of an unbiased estimator is defined by:

$$eff(\hat{a}) = \frac{\frac{1}{I_{\vec{R}}(a)}}{var(\hat{a})} \leq 1$$

The lower bound on the information inequality is achieved if the correlation between $\mathbb{V}(\vec{R};a)$ and $\hat{a}(\vec{R})$ is $+1$ or -1, i.e they are perfectly correlated.
In this case, the score can be expressed as a linear function of the estimate:

$$\mathbb{V} \triangleq f_1'(a).\hat{a}(\vec{R}) + f_2'(a)$$

where $\mathbb{V} = \frac{\delta}{\delta a} \ln p(\vec{R};a)$ and both $f_1'(a)$ and $f_2'(a)$ are invariant in $\vec{R}$ and non-random.
Integrate both sides with respect to the variable a:

$$\ln p(\vec{R};a) = f_1(a).\hat{a}(\vec{R}) + f_2(a) + \underbrace{f_3(\vec{R})}_{\textbf{const. w.r.t. a}}$$

Thus the channel pdf can be expressed via the following model:

$$p(\vec{R};a) = \exp\left[f_1(a).\hat{a}(\vec{R}) + f_2(a) + f_3(\vec{R})\right]$$

that belongs to the exponential family of distributions. For example: Consider the Gaussian pdf

$$p(r) = N(\mu, \sigma^2) : \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(r-\mu)^2}{2\sigma^2}\right]$$

Fix σ and let $a = \mu$.

$$\begin{aligned} \therefore f_2(a) &= \tfrac{-\mu^2}{2\sigma^2} - \ln\sqrt{2\pi}\sigma \\ f_3(r) &= \tfrac{-r^2}{2\sigma^2} \\ f_1(a).f_n(r) &= \tfrac{\mu}{\sigma}.\tfrac{r}{\sigma} \end{aligned}$$

Conclusion If an efficient estimator exists, then the channel pdf has to belong to the exponential family of distribution.

Relation to the ML estimator

We know that the ML estimator is achieved when $p(\vec{R};a)$ attains its maximum, i.e.

$$\left.\frac{\delta}{\delta a} p(\vec{R};a)\right|_{a=\hat{a}_{ML}} = 0$$

or, since ln is a monotone transformation,

$$\left.\frac{\delta}{\delta a} \ln p(\vec{R};a)\right|_{a=\hat{a}_{ML}} = 0$$

If $\hat{a}(\vec{R})$ is an efficient estimator, we showed that

$$\mathbb{V} = f_1'(a).\hat{a}(\vec{R}) + f_2'(a)$$

i.e. $\mathbb{V}$ is a linear function of $\hat{a}(\vec{R})$.
Evaluate the above at $a = \hat{a}_{ML}$:

$$\mathbb{V}\Big|_{a=\hat{a}_{ML}} = f_1'(\hat{a}_{ML}).\hat{a}(\vec{R}) + f_2'(\hat{a}_{ML}) = 0$$

$$\Rightarrow \hat{a}(\vec{R}) = -\frac{f_2'(\hat{a}_{ML})}{f_1'(\hat{a}_{ML})} \tag{1}$$

We also note that the mean value of the score is zero, i.e.

$$\begin{aligned} E(\mathbb{V}) &= E\big[f_1'(a).\hat{a}(\vec{R}) + f_2'(a)\big] = 0 \\ &= f_1'(a)E\big[\hat{a}(\vec{R}\big] + f_2'(a) = 0 \\ \therefore f_1'(a).a + f_2'(a) &= 0 \\ \therefore a &= -\tfrac{f_2'(\hat{a})}{f_1'(\hat{a})} \end{aligned}$$

For $a = \hat{a}_{ML}$, this yields

$$\Rightarrow \hat{a}_{ML} = -\frac{f_2'(\hat{a}_{ML})}{f_1'(\hat{a}_{ML})} \tag{2}$$

Comparing 1 and 2, we obtain:

$$\hat{a}(\vec{R}) = \hat{a}_{ML}$$

Thus the ML estimate is an efficient estimate if one exists$\equiv$ pdf $p(\vec{R}; a)$ belongs to the exponential family.
Note: If $p(\vec{R}; a)$ is not from the exponential family, then the ML estimate cannot be efficient; in fact, there is no efficient estimate in that case.

Multiparameter estimation

$\vec{A}_{K\times 1}$ is the unknown parameter and $\vec{R}$ is the observation vector.

1. Random vector $\vec{A}$, i.e. $p(\vec{A})$ is given.
   Define the estimator error as:

   $$\vec{\epsilon}_{\hat{A}} = A - \hat{\vec{A}}(\vec{R})$$

   (a) MMSE estimator:

   $$\underbrace{min}_{\hat{A}} E\big[\vec{\epsilon}_{\hat{A}}^{\,T}.\vec{\epsilon}_{\hat{A}}\big] = \underbrace{min}_{\hat{A}}\,[\sum_{i=1}^{K}\{a_i - \hat{a}_i(\vec{R})\}^2]$$

   The solution is the conditional mean of the parameter $\hat{A}_{\vec{MMSE}} = E(\vec{A}|\vec{R})$

   (b) Maximum a posteriori (MAP) estimate is located at the global maximum of $p(\vec{A}|\vec{R})$.

   $$\bigtriangledown_{\vec{A}} \quad p(\vec{A}|\vec{R})\Big|_{\vec{A}=\hat{A}_{MAP}} = 0$$

2. Non random parameter estimation (a priori pdf $p(\vec{A})$ is not available).
   ML estimator is given by:

   $$\bigtriangledown_{\vec{A}} \quad p(\vec{R}; \vec{A})\Big|_{\vec{A}=\hat{A}_{ML}} = 0$$

Cramer Rao Bound for multiparameter estimation

Definition: An unbiased estimator has an expected value that is equal to the unknown parameter:

$$E_{\vec{R}}[\hat{\vec{A}}(\vec{R})] = \vec{A}$$

Cramer Rao bound on the variance of an unbiased estimate of $\vec{A}$:
The variance of the estimate of a_i, i.e. $\hat{a}_i(\vec{R})$ is bounded by

$$\sigma_i^2 \geq J^i$$

where J^i is the ith diagonal element of $\vec{J}^{-1}$ and $\vec{J}$ is known as the Fisher information matrix whose (i,j)th element is defined by the following:

$$E\Big[\frac{\delta}{\delta a_i}\ln p(\vec{R};\vec{A}).\frac{\delta}{\delta a_j}\ln p(\vec{R};\vec{A})\Big] = -E\Big[\frac{\delta^2}{\delta a_i \delta a_j}\ln p(\vec{R};\vec{A})\Big]$$

Proof: Define the ith score via

$$\mathbb{V}_i \triangleq \frac{\delta}{\delta a_i}\ln p(\vec{R};\vec{A})$$

where $i = 1, 2, \cdots, K$. The (i,j)th element of the Fisher information matrix is given by:

$$J^{ij} = E(\mathbb{V}_i \mathbb{V}_j)$$

Also, the error for the ith element is defined by:

$$\epsilon_i(\vec{R}) \triangleq \hat{a}_i(\vec{R}) - a_i$$

Define the following vector:

$$\vec{I}_{(K+1)\times 1} \triangleq \begin{bmatrix} \epsilon_i \\ \mathbb{V}_1 \\ \mathbb{V}_2 \\ \vdots \\ \mathbb{V}_K \end{bmatrix} = \begin{bmatrix} \epsilon_i \\ \frac{\delta}{\delta a_1}\ln p(\vec{R};\vec{A}) \\ \frac{\delta}{\delta a_2}\ln p(\vec{R};\vec{A}) \\ \vdots \\ \frac{\delta}{\delta a_K}\ln p(\vec{R};\vec{A}) \end{bmatrix}$$

We know that:

$$\begin{aligned} E(\mathbb{V}_m) &= \int_{\mathbb{Z}} \left[\frac{\delta}{\delta a_m}\ln p(\vec{R};\vec{A})\right] p(\vec{R};\vec{A}) d\vec{R} \\ &= \int_{\mathbb{Z}} \left[\frac{\frac{\delta}{\delta a_m} p(\vec{R};\vec{A})}{p(\vec{R};\vec{A})}\right] .p(\vec{R};\vec{A}) d\vec{R} \\ &= \frac{\delta}{\delta a_m}\int_{\mathbb{Z}} \underbrace{p(\vec{R};\vec{A})}_{=1} d\vec{R} \\ &= 0 \\ \Rightarrow E(\mathbb{V}_m) &= 0 \end{aligned}$$

for $m = 1, 2, \cdots, K$.
If $\hat{\vec{A}}(\vec{R})$ is an unbiased estimate, then $E(\epsilon_i) = 0$ for $i = 1, 2, \cdots, K$.
Thus for the vector $\vec{I}$, we have:

$$E\begin{bmatrix} \epsilon_i \\ \mathbb{V}_1 \\ \mathbb{V}_2 \\ \vdots \\ \mathbb{V}_K \end{bmatrix} = E(\vec{I}) = \vec{0}$$

Construct the covariance matrix of $\vec{I}$:

$$
\begin{aligned}
Q &= E[\vec{I}.\vec{I}^T] \\
&= E\left\{ \begin{bmatrix} \epsilon_i \\ \mathbb{V}_1 \\ \mathbb{V}_2 \\ \vdots \\ \mathbb{V}_K \end{bmatrix} .[\ \epsilon_i \quad \mathbb{V}_1 \quad \cdots \quad \mathbb{V}_K\ ] \right\}_{(K+1)\times(K+1)} \\
&= \begin{bmatrix} E(\epsilon_i^2) & E(\epsilon_i \mathbb{V}_1) & \cdots & E(\epsilon_i \mathbb{V}_K) \\ & \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots\cdots \\ \vdots & \vdots & \{E(\mathbb{V}_n \mathbb{V}_m)\} & \vdots \\ & \vdots & & \vdots \\ E(\epsilon_i \mathbb{V}_K) & \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots\cdots \end{bmatrix}
\end{aligned}
$$

which gives:

$$
Q = \begin{bmatrix} \sigma_i^2 & q_{1,2} & \cdots & q_{1,(K+1)} \\ & \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots\cdots \\ \vdots & \vdots & \vec{J} = \{E(\mathbb{V}_n \mathbb{V}_m)\} & \vdots \\ & \vdots & & \vdots \\ q_{(K+1),1} & \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots\cdots \end{bmatrix}
$$

where $q_{i,j} = q_{j,i} = E(\epsilon_i \mathbb{V}_{j-1})$ for $j = 2, 3, \cdots, (K+1)$, and $\vec{J}$ is called the Fisher Information matrix whose (n,m)th component is $E(\mathbb{V}_n)\mathbb{V}_m$.

$$
\begin{aligned}
q_{i,(j+1)} &= E(\epsilon_i, \mathbb{V}_j) \qquad\qquad \forall j = 1, 2, ..., K \\
&= E[\epsilon_i \tfrac{\delta}{\delta a_j} \ln p(\vec{R}; \vec{A})] \\
&= E\left[\frac{\epsilon_i \frac{\delta}{\delta a_j} p(\vec{R}; \vec{A})}{p(\vec{R}; \vec{A})} \right]
\end{aligned}
$$

We use the derivative identity

$$
\mathbf{X}.\mathbf{Y}' = (\mathbf{X}\mathbf{Y})' - \mathbf{X}'\mathbf{Y}
$$

Put $\mathbf{X} = \epsilon_i$, $\mathbf{Y} = p(\vec{R};\vec{A})$.

$$\begin{aligned}
\Rightarrow \quad & = \epsilon_i \tfrac{\delta}{\delta a_j} p(\vec{R};\vec{A}) \\
& = \tfrac{\delta}{\delta a_j}[\epsilon_i p(\vec{R};\vec{A})] - \tfrac{\delta}{\delta a_j}\epsilon_i p(\vec{R};\vec{A}) \\
& = E\left[\frac{\frac{\delta}{\delta a_j}[\epsilon_i p(\vec{R};\vec{A})]}{p(\vec{R};\vec{A})}\right] - E\left[\frac{\frac{\delta}{\delta a_j}\epsilon_i p(\vec{R};\vec{A})}{p(\vec{R};\vec{A})}\right] \\
& = \int_{\mathbb{Z}} \left[\frac{\frac{\delta}{\delta a_j}[\epsilon_i p(\vec{R};\vec{A})]}{p(\vec{R};\vec{A})}\right].p(\vec{R};\vec{A})d\vec{R} - \int_{\mathbb{Z}} \epsilon_i p(\vec{R};\vec{A})d\vec{R} \\
& = \tfrac{\delta}{\delta a_j} \underbrace{\int_{\mathbb{Z}} \epsilon_i p(\vec{R};\vec{A})d\vec{R}}_{=0 \text{ since estimate is unbiased}} - \int_{\mathbb{Z}} \tfrac{\delta}{\delta a_j}[\hat{a}_i - a_i] p(\vec{R};\vec{A})d\vec{R} \\
& = \tfrac{\delta}{\delta a_j} 0 - \int_{\mathbb{Z}} (-\delta_{ij}) p(\vec{R};\vec{A})d\vec{R} \\
& = \delta_{ij} \int_{\mathbb{Z}} p(\vec{R};\vec{A})d\vec{R} = \delta_{ij}
\end{aligned}$$

Thus the covariance matrix becomes:

$$\vec{Q} = \left[\begin{array}{cccccccc}
 & & & & (i+1)st & & & \\
 & \sigma_i^2 & 0 & 0 & \cdots\cdots & 1 & 0 & \cdots & 0 \\
 & 0 & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\
 & 0 & \vdots & & & | & & & \vdots \\
 & \vdots & \vdots & & & | & & & \vdots \\
(i+1)st & 1 & \vdots -- & --- & --- & J_{ii} & & & \vdots \\
 & \vdots & \vdots & & & & & & \vdots \\
 & 0 & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots
\end{array}\right]$$

We know that the determinant of a covariance matrix is always non negative; thus

$$|\vec{Q}| \geq 0$$

However

$$\begin{aligned}
|\vec{Q}| \quad & = \sigma_i^2|\vec{J}| + 0 + 0 + \cdots + (-1)^i 1.cofactor(J_{ii}) + 0 + \cdots + 0 \geq 0 \\
\Rightarrow \sigma_i^2 \quad & \geq \tfrac{-(-1)^i 1.cofactor(J_{ii})}{|\vec{J}|} \\
& \triangleq J^{ii}
\end{aligned}$$

where J^{ii} is the (i,i)th element of $\vec{J}^{-1}$. Therefore: $\boxed{\text{CR bound}: \ \sigma_i^2 \geq J^{ii}}$

The bound is achieved when $|\vec{Q}| = 0$.
This occurs if the components of $\vec{I}$ are linearly dependent. i.e.

$$\begin{aligned}
\epsilon_i \quad & = \textstyle\sum_{j=1}^{K} f'_{1,ij}(\vec{A})\mathbb{V}_j + f'_{2i}(\vec{A}) \\
& = \textstyle\sum_{j=1}^{K} f'_{1,ij}(\vec{A}) \tfrac{\delta}{\delta a_j} \ln p(\vec{R};\vec{A}) + f_{2i}(\vec{A})
\end{aligned}$$

For the random case of $\vec{A}$

$$\ln p(\vec{R};\vec{A}) = \ln p(\vec{R}|\vec{A}) + \ln p(\vec{A})$$

Thus there will be two Fisher information matrices viz. $\vec{J}_{\vec{R}|\vec{A}}$ and $\vec{J}_{\vec{A}}$.
The same procedure as the non-random case will be followed to show that

$$\sigma_i^2 \geq J_T^{ii}$$

where J_T^{ii} is the (i,i)th element of J_T^{-1} and

$$J_T = \vec{J}_{\vec{R}|\vec{A}} + \vec{J}_{\vec{A}}$$

is the total Fisher information matrix.

Composite Hypotheses

$$H_i: \quad p(\vec{R}|H_i;\vec{\theta_i})$$

where $\vec{\theta_i}$ depends on H_i and is unknown.

Objective: We are interested in deciding among H_i without actually caring about $\vec{\theta_i}$.
example: Data communication via FSK.

$$\begin{aligned} H_0: &\quad r(t) = A\cos(\omega_0 t + \theta) + n(t) \\ H_1: &\quad r(t) = A\cos(\omega_1 t + \theta) + n(t) \end{aligned}$$

where θ (phase) is unknown and we are interested in deciding H_0 or H_1 (or ω_0 or ω_1). θ is called the unwanted parameter.

Case 1: Random parameter:
A priori pdf $p(\theta_i|H_i)$ is known.
Generalized likelihood ratio test (GLRT) is constructed via

$$\begin{aligned} \Lambda_i(\vec{R}) &= \frac{\int_{\theta_i} p(\vec{R}|H_i,\theta_i)p(\theta_i|H_i)d\theta_i}{\int_{\theta_0} p(\vec{R}|H_0,\theta_0)p(\theta_0|H_0)d\theta_0} \\ &= \frac{p(\vec{R}|H_i)}{p(\vec{R}|H_0)} \end{aligned}$$

Case 2: Non-random parameter:
A priori pdf is unknown. To construct the LRT, use the ML estimate of θ_i.

$$\Lambda_i(\vec{R}) = \frac{max_{\theta_i} p(\vec{R}|H_i,\theta_i)}{max_{\theta_0} p(\vec{R}|H_0,\theta_0)}$$

General Gaussian problem

- For a detection problem $p(\vec{R}|H_i)$ is a multivariate Gaussian.
For an estimation problem $p(\vec{R}|\vec{A})$ is a multivariate Gaussian.
Let M be the mean of $\vec{R}$ and $\vec{\Lambda}$ be its covariance matrix.

$$\begin{aligned} \vec{M}_{N\times 1} &\triangleq E(\vec{R}) \\ \vec{\Lambda}_{N\times N} &\triangleq E\big[(\vec{R}-\vec{M})(\vec{R}-\vec{M})^T\big] \end{aligned}$$

where $\vec{R}$ is a multivariate normal distribution and its pdf is given by:

$$p(\vec{R}) = \frac{1}{(\sqrt{2\pi})^N |\vec{\Lambda}|^{1/2}} \exp\big[-\frac{1}{2}(\vec{R}-\vec{M})\vec{\Lambda}^{-1}(\vec{R}-\vec{M})^T \big]$$

Any linear transformation of a Gaussian vector is also Gaussian.

$$\vec{S}_{L\times 1} = \vec{A}_{L\times N}\vec{R}_{N\times 1}$$

where $L \leq N$ and $\vec{A}$ is a deterministic matrix with rank L. In this case, S is also multivariate Gaussian with

$$\begin{aligned} E(\vec{S}) &= \vec{A}\vec{M} \\ cov(\vec{S}) &= \vec{A}.\vec{\Lambda}.\vec{A}^T \\ &\triangleq \vec{\Lambda}_S \end{aligned}$$

<u>Objective:</u> Identify a linear transformation of $\vec{R}$ that yields uncorrelated (independent) r.v.'s

$$\vec{R}"_{N\times 1} = \vec{W}_{N\times N}\vec{R}_{N\times 1}$$

<u>Solution:</u>
We write:

$$\vec{W} = \begin{bmatrix} \vec{\Phi_1}^T \\ \vdots \\ \vec{\Phi_N}^T \end{bmatrix}$$

For $\vec{R}"$ to have uncorrelated components, the $cov(\vec{R}")$ should be a diagonal matrix.

$$\begin{aligned} cov(\vec{R}") &= E\big[(\vec{R}" - \vec{M}")(\vec{R}" - \vec{M}")^T\big] \\ &\triangleq \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & \sigma_N^2 \end{bmatrix} \end{aligned}$$

where

$$\vec{M}" = E(\vec{R}") = \vec{W}\vec{M}$$

Substitute for $\vec{R}"$ in the above covariance:

$$\begin{aligned} cov(\vec{R}") &= E\big[\vec{W}(\vec{R}-\vec{M})(\vec{R}-\vec{M})^T\vec{W}^T\big] \\ &= \vec{W}\big[(\vec{R}-\vec{M})(\vec{R}-\vec{M})^T\big]\vec{W}^T \\ &= \vec{W}\vec{\Lambda}\vec{W}^T \\ &\triangleq \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & \sigma_N^2 \end{bmatrix} \end{aligned}$$

The $(i,j)th$ element in $cov(\vec{R}")$ is

$$\vec{\Phi}_i^T \vec{\Lambda} \vec{\Phi}_j = \begin{cases} \sigma_i^2 & ; i = j \\ 0 & ; i \neq j \end{cases}$$

For above to be true, the Φ_i's should be chosen to be the eigenvectors of the $\vec{\Lambda}$ matrix. The σ_i^2s are the eigen values of $\vec{\Lambda}$. These eigenvectors and eigenvalues are the solutions of

$$|\vec{\Lambda} - \sigma^2 \vec{I}| = 0$$

If we need, equal variances, e.g. unit variances, we can define the following "scaled" transformation of $\vec{R}"$.

$$\vec{R}' = \vec{\Sigma}^{-1} \vec{R}"$$

where

$$\vec{\Sigma}^{-1} = \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & 1/\sigma_N \end{bmatrix}$$

or

$$\vec{\Sigma} = \frac{1}{\prod_{i=1}^{N} \sigma_i} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & \sigma_N \end{bmatrix}$$

Covariance of $\vec{R}'$ is $cov(\vec{R}') = \vec{I}_{N \times N}$.

$$\begin{aligned} cov(\vec{R}') &= \vec{\Sigma}^{-1} cov(\vec{R}") \vec{\Sigma}^{-1T} \\ &= \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & 1/\sigma_N \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & \sigma_N^2 \end{bmatrix} \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & 1/\sigma_N \end{bmatrix} \\ &= \vec{I} \end{aligned}$$

The overall transformation is:

$$\vec{R}' = \underbrace{\vec{\Sigma}^{-1} \vec{W}}_{\vec{H}:\text{transformation matrix}} \vec{R}$$

$$\Rightarrow \quad E(\vec{R}') = \vec{H}.\vec{M}$$

$$cov(\vec{R}') = \vec{I}$$

Note that if $\vec{R}$ is a sufficient statistic, so is $\vec{R}'$ since there is no loss in dimensionality. In general, we can treat a Gaussian detection/estimation problem as passing the received vector through a linear transformation

$$\vec{H} = \vec{\Sigma}^{-1} \vec{W}$$

and then designing a receiver based on $\vec{R}'$.