Multiple parameter Estimation

Consider

\[ c(a, \hat{a}) = c(\hat{a} - a) = c(a_e(R)) \]

For MS Criterion

\[ c(a_e(R)) = \sum_{i=1}^{K} a_i^2(R) = a_e^T(R)a_e(R). \]

Jalew

\[ R_{ms} = \iint c(a_{e(R)}) \frac{P(A|R)}{P_e(R)} dR dA \]

\[ = \int_{-\infty}^{\infty} P_r(R) \left[ \int_{-\infty}^{\infty} c(a_{e(R)}) P(A|R) dA \right] dR \]

\[ = \int_{-\infty}^{\infty} P_r(R) \left[ \int_{-\infty}^{\infty} \sum_{i=1}^{K} (\hat{a}_i - A_i)^2 P(A|R) dA \right] dR \]

minimizing \( R_{ms} \) = minimizing the inner integral \( I \)

for each \( R \). Since the terms in the sum are positive, we minimize them separately.
\[ I = \int_{-\infty}^{\infty} (\hat{A}_i - A_i)^2 P_{\text{air}}(A|B) dA + \ldots + \int_{-\infty}^{\infty} (\hat{A}_k - A_k)^2 P_{\text{air}}(A|B) dA. \]

\[ I_1 \]

\[ \left\{ \begin{align*}
\frac{\partial I_1}{\partial \hat{A}_i} &= 0 \\
\frac{\partial I_k}{\partial \hat{A}_k} &= 0
\end{align*} \right. \]

\[ \hat{A}_i = \int_{-\infty}^{\infty} A_i P_{\text{air}}(A|B) dA \]

\[ \hat{A}_k = \int_{-\infty}^{\infty} A_k P_{\text{air}}(A|B) dA. \]

\[ A_{\text{ms}} = \int_{-\infty}^{\infty} A P_{\text{air}}(A|B) dA \]

For MAP estimation, we should have:

\[ \frac{\partial \ln P_{\text{air}}(A|B)}{\partial A_i} \bigg|_{A = \hat{A}_{\text{map}}} = 0, \quad i = 1, 2, \ldots, k \]
\[ \nabla_A \left[ \ln P_{AIR}(A|B) \right] \bigg|_{A = \hat{A}_{MAP}(B)} = 0. \]

where

\[ \nabla_A = \left( \frac{\partial}{\partial A_1}, \ldots, \frac{\partial}{\partial A_k} \right)^T \]

Similarly, for ML estimates we must find \( A = \hat{A}_{ML}(B) \) that maximizes \( P_{R1A}(R|IA) \).

\[ \nabla_A \left[ \ln P_{R1A}(R|IA) \right] \bigg|_{A = \hat{A}_{ML}(B)} = 0. \]

In both cases we must verify that we have absolute maximum.
For ML estimation we must find the value of $A$ that maximizes $p_{\text{ML}}(A|R)$. If the maximum is interior and $\frac{\partial \ln p_{\text{ML}}(A|R)}{\partial A_i}$ exists at the maximum then a necessary condition is obtained from the MAP equations. By analogy with (137) we take the logarithm of $p_{\text{ML}}(A|R)$, differentiate with respect to each parameter $A_i$, $i = 1, 2, \ldots, K$, and set the result equal to zero. This gives a set of $K$ simultaneous equations:

$$\frac{\partial \ln p_{\text{ML}}(A|R)}{\partial A_i} = 0, \quad i = 1, 2, \ldots, K. \quad (238)$$

We can write (238) in a more compact manner by defining a partial derivative matrix operator

$$\nabla_A \triangleq \begin{bmatrix}
\frac{\partial}{\partial A_1} \\
\frac{\partial}{\partial A_2} \\
\vdots \\
\frac{\partial}{\partial A_K}
\end{bmatrix}. \quad (239)$$

This operator can be applied only to $1 \times m$ matrices; for example,

$$\nabla_A G = \begin{bmatrix}
\frac{\partial G_1}{\partial A_1} & \frac{\partial G_2}{\partial A_1} & \cdots & \frac{\partial G_m}{\partial A_1} \\
\frac{\partial G_1}{\partial A_2} & \frac{\partial G_2}{\partial A_2} & \cdots & \frac{\partial G_m}{\partial A_2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial G_1}{\partial A_K} & \frac{\partial G_2}{\partial A_K} & \cdots & \frac{\partial G_m}{\partial A_K}
\end{bmatrix}. \quad (240)$$

Several useful properties of $\nabla_A$ are developed in Problems 2.4.27–28. In our case (238) becomes a single vector equation,

$$\nabla_A \ln p_{\text{ML}}(A|R) = 0. \quad (241)$$

Similarly, for ML estimates we must find the value of $A$ that maximizes $p_{\text{ML}}(R|A)$. If the maximum is interior and $\frac{\partial \ln p_{\text{ML}}(R|A)}{\partial A_i}$ exists at the maximum then a necessary condition is obtained from the likelihood equations:

$$\nabla_A [\ln p_{\text{ML}}(R|A)]_{A = \hat{A}_{\text{MAP}}} = 0. \quad (242)$$

In both cases we must verify that we have the absolute maximum.

**Measures of Error.** For nonrandom variables the first measure of interest is the bias. Now the bias is a vector,

$$B(A) \triangleq E[\hat{A}(R)] - A. \quad (243)$$

If each component of the bias vector is zero for every $A$, we say that the estimate is unbiased.

In the single parameter case a rough measure of the spread of the error was given by the variance of the estimate. In the special case in which $a(R)$ was Gaussian this provided a complete description:

$$p_{a}(A) = \frac{1}{\sqrt{2\pi}a_x} \exp \left(-\frac{A^2}{2a_x^2}\right). \quad (244)$$

For a vector variable the quantity analogous to the variance is the covariance matrix

$$\mathbb{E}[(\hat{a} - \bar{a})(\hat{a}^T - \bar{a}^T)] \triangleq \Lambda_x, \quad (245)$$

where

$$\bar{a} \triangleq E(a) = B(A). \quad (246)$$

The best way to determine how the covariance matrix provides a measure of spread is to consider the special case in which the $a_i$ are jointly Gaussian. For algebraic simplicity we let $E(a) = 0$. The joint probability density for a set of $K$ jointly Gaussian variables is

$$p_{a}(A) = \left(\frac{\kappa K^m}{2\pi}\right)^{\frac{1}{2}} \exp \left(-\frac{1}{2}A^T \Lambda_x^{-1} A\right). \quad (247)$$

(e.g., p. 151 in Davenport and Root [1]).

The probability density for $K = 2$ is shown in Fig. 2.24a. In Figs. 2.24b, c we have shown the equal-probability contours of two typical densities. From (247) we observe that the equal-height contours are defined by the relation

$$A_i^T \Lambda_x^{-1} A_i = C^2, \quad (248)$$

which is the equation for an ellipse when $K = 2$. The ellipses move out monotonically with increasing $C$. They also have the interesting property that the probability of being inside the ellipse is only a function of $C^2$.

**Property.** For $K = 2$, the probability that the error vector lies inside an ellipse whose equation is

$$A_i^T \Lambda_x^{-1} A_i = C^2, \quad (249)$$

is

$$P = 1 - \exp \left(-\frac{C^2}{2}\right). \quad (250)$$

**Proof.** The area inside the ellipse defined by (249) is

$$A = \left| A_i \right|^{\frac{1}{2}} \frac{1}{2\pi C} \, dC. \quad (251)$$

The differential area between ellipses corresponding to $C$ and $C + dC$ respectively is

$$dA = \left| A_i \right|^{\frac{1}{2}} 2\pi C \, dC. \quad (252)$$
For this reason the ellipses described by (248) are referred to as concentration ellipses because they provide a measure of the concentration of the density.

A similar result holds for arbitrary $K$. Now, (248) describes an ellipsoid. Here the differential volume† in $K$-dimensional space is

$$dv = |A_e|^{1/2} \frac{\pi^{K/2}}{\Gamma(K/2 + 1)} KC^{K-1} dC. \quad (255)$$

The value of the probability density on the ellipsoid is

$$[(C\pi)^{K/2} |A_e|^{1/2}]^{-1} \exp \left( -\frac{C^2}{2} \right). \quad (256)$$

Therefore

$$1 - P = \frac{K}{(2\pi)^{K/2} \Gamma(K/2 + 1)} \int_0^\infty X^{K-1} e^{-X^{2/2}} dX, \quad (257)$$

which is the desired result. We refer to these ellipsoids as concentration ellipsoids.

When the probability density of the error is not Gaussian, the concentration ellipsoid no longer specifies a unique probability. This is directly analogous to the one-dimensional case in which the variance of a non-Gaussian zero-mean random variable does not determine the probability density. We can still interpret the concentration ellipsoid as a rough measure of the spread of the errors. When the concentration ellipsoids of a given density lie wholly outside the concentration ellipsoids of a second density, we say that the second density is more concentrated than the first. With this motivation, we derive some properties and bounds pertaining to concentration ellipsoids.

**Bounds on Estimation Errors; Nonrandom Variables.** In this section we derive two bounds. The first relates to the variance of an individual error; the second relates to the concentration ellipsoid.

**Property I.** Consider any unbiased estimate of $A_i$. Then

$$\sigma_i^2 = \operatorname{Var} [\hat{a}(R) - A_i] \leq J_i^T, \quad (258)$$

where $J_i$ is the $i$th element in the $K \times K$ square matrix $J$. The elements in $J$ are

$$J_{ij} = E \left[ \frac{\partial}{\partial A_i} \ln p_{\text{est}}(R|A) \frac{\partial}{\partial A_j} \ln p_{\text{est}}(R|A) \right]$$

$$= -E \left[ \frac{\partial^2 \ln p_{\text{est}}(R|A)}{\partial A_i \partial A_j} \right]. \quad (259)$$

† e.g., Cramér [9], p. 120, or Sommerfeld [32].
1. Let $\hat{A}_i(B)$ represent any unbiased estimate of $A_i$. Then

$$e_i^2 = \text{Var} [\hat{A}_i(B) - A_i] \geq J^{ii},$$

where $J^{ii}$ is the $ii$th element in the $K \times K$ square matrix $J^{-1}$. The elements in $J$ are

$$J_{ij} = \mathbb{E} \left[ \frac{\text{coln} P_{xi} (B|A)}{\text{coln} P_{xi} (B|A)} \right]$$

$$= -\mathbb{E} \left[ \frac{\text{coln} P_{xi} (B|A)}{\text{coln} P_{xi} (B|A)} \right].$$

The $J$ matrix is commonly called Fisher's information matrix. The equality in (4) holds iff

$$\hat{A}_i(B) - A_i = \sum_{j=1}^{K} K_{ij}(A) \frac{\text{coln} P_{xi} (B|A)}{\text{coln} P_{xi} (B|A)}.$$
**Proof**: Because $\hat{A}_i(R)$ is unbiased, $(\hat{A}_i(R) = A_i)$

$$\int_{-\infty}^{\infty} [\hat{A}_i(R) - A_i] P_{1a}(R|IA) \, dR = 0, \quad j = 1, k$$

or

$$\int_{-\infty}^{\infty} \hat{A}_i(R) P_{1a}(R|IA) \, dR = \sum_{c_{IA}} A_i \quad i = 1, k$$

$$\implies \int_{-\infty}^{\infty} \hat{A}_i(R) \left( \frac{P_{1a}(R|IA)}{P_{1a}(R|IA)} \right) \, dR = \Sigma_{i,j}$$

Let $i = 1$. Define a $k+1$ vector

$$X = \begin{bmatrix}
\hat{A}_1(R) - A_1 \\
\frac{\ln P_{1a}(R|IA)}{c_{IA}} \\
\vdots \\
\frac{\ln P_{1a}(R|IA)}{c_{IA}} \\
\end{bmatrix}
\begin{cases}
K+1 \text{ elements}
\end{cases}$$
The covariance matrix is

\[ E \{XX^T \} = E \left\{ \begin{bmatrix} \hat{\alpha}_1(B) - A_1 \\ \frac{\ln P_{11}(BIA)}{\sigma A_1} \\ \vdots \\ \frac{\ln P_{1k}(BIA)}{\sigma A_k} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1(B) - A_1 \\ \frac{\ln P_{11}(BIA)}{\sigma A_1} \\ \vdots \\ \frac{\ln P_{1k}(BIA)}{\sigma A_k} \end{bmatrix} \right\} \]

\[ = \begin{bmatrix} E \{(\hat{\alpha}_1(B) - A_1)^2 \} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \]

\[ = \begin{bmatrix} 1 & J_{11} & \cdots & J_{1k} \\ J_{11} & 1 & \cdots & J_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ J_{k1} & J_{k2} & \cdots & 1 \end{bmatrix} \]

Aside:

Note 1: \[ E \{ (\hat{\alpha}_1(B) - A_1) \frac{\ln P_{11}(BIA)}{\sigma A_1} \} \]

\[ = E \{ \hat{\alpha}_1(B) \frac{\ln P_{11}(BIA)}{\sigma A_1} \} - E \{ A_1 \frac{\ln P_{11}(BIA)}{\sigma A_1} \} \]

But \[ E \{ A_1 \frac{\ln P_{11}(BIA)}{\sigma A_1} \} = A_1 \int \frac{\ln P_{11}(BIA)}{\sigma A_1} P(BIA) dBIA \]

\[ = A_1 \frac{\int P_{11}(BIA) dBIA}{\sigma A_1} = 0 \]
Note 2:

\[ E \left\{ (\mathbf{a}_i^T - \mathbf{A}_i) \frac{\partial \ln P_{110}(\mathbf{RIA})}{\partial \mathbf{A}_2} \right\} = E \left\{ \mathbf{a}_i^T \frac{\partial \ln P_{110}(\mathbf{RIA})}{\partial \mathbf{A}_2} \right\} \]

\[ -\mathbf{A}_i \frac{\partial \ln P_{110}(\mathbf{RIA})}{\partial \mathbf{A}_2} \]

\[ \Rightarrow 0. \]

Proof contd:

Since it is a covariance matrix, it is nonnegative definite, which implies that

\[ \det \{ E(XX^T) \} \geq 0. \]

Note that

\[ \det \left\{ E^{1/2}(XX^T) \right\} = 0^{2} \frac{1}{|J|} - \text{cofactor } J_{11} \geq 0. \]

\[ 0^2 \geq \frac{\text{cofactor } J_{11}}{|J|} = J'' \]

The generalization to \( i \neq 1 \) is simple.

QED
Note 3:

\[ \text{det} \{ E \{ x \times T \} \} = \varepsilon_1^2 \text{ Cofactor } \varepsilon_1^2 - 1 \text{ Cofactor } 1 \]

\[ \text{Cofactor } 1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \\ J_{K1} & \cdots & J_{K2} & J_{KK} \end{pmatrix} \]

\[ = 1 \text{ Cofactor } 1 + 0 + 0 \cdots = \text{ Cofactor } J_{11} \]

\[ \begin{pmatrix} J_{11} & \cdots & J_{1K} \\ \vdots & \ddots & \vdots \\ J_{K1} & \cdots & J_{KK} \end{pmatrix} \]

\[ \text{Cofactor } J_{11} \]

\[ \therefore \text{det} \{ E \{ x \times T \} \} = \varepsilon_1^2 |J| - \text{Cofactor } J_{11}. \]