

APPENDIX A: PROBABILITY SPACE AND RANDOM VARIABLES

We shall consider *sets of elements or points*. The nature of the points needs not be defined, they may represent elementary events, real values, etc. A *set* is an aggregate of such points. The following operations on two sets A and B are useful:

$A \cup B$, union of A and B , is the set of all points in either A or B

$A \cap B$, intersection of A and B , is the set of all points that belong to both A and B

$A - B$, difference of A minus B , is the set of all points in A that are not in B

DEFINITION A.1. If Ω is a given set, then a σ -algebra F on Ω is a family F of subsets of Ω with the following properties:

(i) $\emptyset \in F$, where \emptyset is the empty set

(ii) $F \in F \Rightarrow F^C \in F$, where $F^C = \Omega \setminus F$ is the complement of F in Ω

(iii) $A_1, A_2, \dots \in F \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in F$

The pair (Ω, F) is called a measurable space. A probability measure P on a measurable space (Ω, F) is a function $P: F \rightarrow [0,1]$ such that

(a) $P(\Omega) = 1$

(b) $0 \leq P(A_i) \leq 1$ for all sets $A_i \in F$

(c) if $A_i \cap A_j = \emptyset$ ($i \neq j$), then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

The triple (Ω, F, P) is called a probability space.

In a probability context the sets are called events, and we use the interpretation

$P(F)$ = "the probability that the event F occurs".

An important σ -algebra used to define random variables is the Borel σ -algebra on \mathfrak{R}^1 where \mathfrak{R}^1 is the real line. The Borel σ -algebra on \mathfrak{R}^1 , noted B^1 , is the smallest σ -algebra containing the collection I of all the open intervals of \mathfrak{R}^1 of the form $(a, b) = \{\chi: a < \chi < b\}$, $(-\infty < a \leq b < \infty)$, namely

$$B^1 = \bigcap \{H: H \sigma\text{-algebra of } \mathfrak{R}^1, I \subset H\}.$$

It can be shown that the Borel σ -algebra B^1 does indeed exist. The elements $B \in B^1$ are called *Borel sets*, and they include all real intervals; open, closed, semiclosed, finite or infinite.

If (Ω, F, P) is a given probability space, then a function $y: \Omega \rightarrow \mathfrak{R}^1$ is called *F*-measurable if

$$y^{-1}(U) := \{\omega \in \Omega: y(\omega) \in U\} \in F.$$

On the strength of the notions introduced we define the random variable as follow

DEFINITION A.2: Let (Ω, F, P) be a given probability space, and (\mathfrak{R}^1, B^1) be a measurable space, where B^1 is a σ -algebra of Borel sets on the real line \mathfrak{R}^1 . A (real valued) random variable $x(\omega)$, where $\omega \in \Omega$ are elementary events, is a *F* measurable mapping from (Ω, F) to (\mathfrak{R}^1, B^1) , i.e. $x: \Omega \rightarrow \mathfrak{R}^1$ and

$$\forall B \in B^1, \quad x^{-1}(B) := \{\omega \in \Omega: x(\omega) \in B\} \in F$$

Every random variable induces a probability measure μ_x on (\mathfrak{R}^1, B^1) , defined by

$$\mu_x(B) = P(x^{-1}(B)) = P[x = x(\omega) \in B^I]$$

From the probability measure μ_x we define the cumulative distribution function of the random variable x

$$F_x(\chi) = \mu_x(I_\chi := \{\chi' : \chi' \leq \chi\}) = P[x \leq \chi]$$

APPENDIX B: SOLUTION OF THE ENTROPY MAXIMIZATION PROBLEM

Let the entropy of the probability distribution $f_x(\mathbf{X}_{map})$ be defined as

$$\varepsilon(f_x) = - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_x(\mathbf{X}_{map})).$$

Consider the constrained maximization problem written in compact form as:

$$\begin{aligned} & \max_{f_x} \varepsilon(f_x) \\ \text{s.t. } & E[g_q] = \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) g_q(\mathbf{X}_{map}), \quad q = 1, \dots, Q \end{aligned}$$

where $g_q(\mathbf{X}_{map})$, $q=1, \dots, Q$ is a suitable set of functions for which the expected values $E[g_q]$ are known. This problem can be solved by introducing the function

$$F[f_x] = - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_x(\mathbf{X}_{map})) - \sum_{q=1}^Q \mu_q \left[\int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) g_q(\mathbf{X}_{map}) - E[g_q] \right]$$

where μ_q are Lagrange coefficients. The constrained maximization problem is equivalent to maximizing $F[f_x]$ with respect to f_x and the Lagrange coefficients μ_q . We can rewrite $F[f_x]$ as

$$F[f_x] = - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_x(\mathbf{X}_{map})) - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \sum_{q=1}^Q \mu_q [g_q(\mathbf{X}_{map}) - E[g_q]]$$

or,

$$F[f_x] = - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_x(\mathbf{X}_{map})) - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_Y(\mathbf{X}_{map})),$$

where

$$f_Y(\mathbf{X}_{map}) = \frac{\exp \sum_{q=1}^Q \mu_q g_q(\mathbf{X}_{map})}{\exp \sum_{q=1}^Q \mu_q E[g_q]}.$$

If it is assumed that $f_Y(\mathbf{X}_{map})$ is a valid probability distribution function, then it can be proven that (see Papoulis, eq 15-139):

$$- \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_x(\mathbf{X}_{map})) \leq - \int_{\mathbf{R}^{m+1}} d\mathbf{X}_{map} f_x(\mathbf{X}_{map}) \ln(f_Y(\mathbf{X}_{map})),$$

leading to $F[f_x] \leq 0$. The maximum is thus reached for $f_x = f_Y$, leading to the solution

$$f_x(\mathbf{X}_{map}) = \frac{\exp \sum_{q=1}^Q \mu_q g_q(\mathbf{X}_{map})}{\exp \sum_{q=1}^Q \mu_q E[g_q]}.$$

It can be shown that $f_Y(\mathbf{X}_{map})$ is a valid distribution, since $f_Y(\mathbf{X}_{map}) > 0$ as it is the ratio of exponentials, and $\int d\mathbf{X}_{map} f_Y(\mathbf{X}_{map}) = 1$ by choice of $g_0(\mathbf{X}_{map}) = 1$.

APPENDIX C. INVERSE OF PARTITIONED MATRICES

C.1 Properties of the inverse of partitioned matrices

Consider the square matrix \mathbf{C} partitioned as follow

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix},$$

where \mathbf{C}_{11} and \mathbf{C}_{22} are square. If \mathbf{C} is non singular than we can write the inverse of \mathbf{C} as

$$\mathbf{A} = \mathbf{C}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

We can write (Searle, 1982, pp 260) the following equation

$$\mathbf{A}^{-1} = \mathbf{C} = \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ -\mathbf{S}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{S}^{-1} \end{bmatrix}, \quad (\text{C.1})$$

where $\mathbf{S} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ is called the Schur complement of \mathbf{A}_{11} in \mathbf{A} . Expanding Eq.

(C.1) we get

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}^{-1} \\ -\mathbf{S}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{S}^{-1} \end{bmatrix}. \quad (\text{C.2})$$

C.2. Property of the first Schur complement

The first Schur complement refers to the following expression

$$\mathbf{S} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}.$$

From Eq. (C.2) we get the following expression for the \mathbf{C}_{22} term

$$\mathbf{C}_{22} = \mathbf{S}^{-1} = \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}, \text{ or equivalently}$$

$$\mathbf{C}_{22}^{-1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}. \quad (\text{C.3})$$

From Eq. (C.2) we get the following expression for the \mathbf{C}_{12} term

$$\mathbf{C}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}^{-1},$$

using $\mathbf{C}_{22} = \mathbf{S}^{-1}$ and re-arranging we get

$$\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = -\mathbf{C}_{12}\mathbf{C}_{22}^{-1} \quad (\text{C.4})$$

C.3. Property of the second Schur complement

The second Schur complement refers to the following expression

$$\mathbf{S}' = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}:$$

Consider $\mathbf{C}' = \begin{bmatrix} \mathbf{C}_{22} & \mathbf{C}_{21} \\ \mathbf{C}_{12} & \mathbf{C}_{11} \end{bmatrix}$ and $\mathbf{A}' = \begin{bmatrix} \mathbf{A}_{22} & \mathbf{A}_{21} \\ \mathbf{A}_{12} & \mathbf{A}_{11} \end{bmatrix}$ where \mathbf{C}_{11} , \mathbf{C}_{12} , ..., \mathbf{A}_{22} are the same as

defined above. Since $\mathbf{A} = \mathbf{C}^{-1}$, we also have $\mathbf{A}' = \mathbf{C}'^{-1}$. Therefore all results obtained for

\mathbf{C}_{11} , \mathbf{C}_{12} , ..., \mathbf{A}_{22} are also valid by swapping the 1 and 2 indices. Thus we consider

$\mathbf{S}' = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ as the Schur complement of \mathbf{A}_{22} in \mathbf{A} , and we have

$$\mathbf{C}_{11}^{-1} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}, \quad (\text{C.5})$$

and,

$$\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = -\mathbf{C}_{21}\mathbf{C}_{11}^{-1}. \quad (\text{C.6})$$

C.4. Properties of the determinant

Using the properties of matrices and determinants we can see from the above equation that

$$|\mathbf{A}| = |\mathbf{A}_{11}||\mathbf{S}| = \frac{|\mathbf{A}_{11}|}{|\mathbf{C}_{22}|}, \quad (\text{C.7})$$

and

$$|\mathbf{A}| = \frac{|\mathbf{A}_{22}|}{|\mathbf{C}_{11}|}. \quad (\text{C.8})$$

APPENDIX D: PROPERTIES OF MULTIVARIATE NORMAL DISTRIBUTIONS

D.1. The Multivariate normal distribution:

Let $\mathbf{x} = [x_1, \dots, x_n]$ be a multivariate normal random vector taking values $\mathbf{X} = [X_1, \dots, X_n]$. The probability distribution of \mathbf{x} is given by

$$f_{\mathbf{x}}(\mathbf{X}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2} \mathbf{Y}^T \mathbf{A} \mathbf{Y}\right],$$

where $\mathbf{m} = E[\mathbf{x}]$ is the vector of expected values, where $\mathbf{Y} = \mathbf{X} - \mathbf{m}$, and where \mathbf{A} is a symmetric positive definite matrix that can be written as $\mathbf{A} = \mathbf{H}^T \mathbf{H}$. The distribution can also be written using the covariance matrix $\mathbf{C} = \mathbf{A}^{-1} = (\mathbf{H}^T \mathbf{H})^{-1} = \mathbf{H}^{-1} \mathbf{H}^{T-1}$ as

$$f_{\mathbf{x}_n}(\mathbf{X}) = \frac{|\mathbf{C}^{-1}|^{1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2} \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y}\right].$$

Note that since \mathbf{A} is symmetric definite positive it can be diagonalized by an orthogonal matrix:

$$\mathbf{\Lambda} = \mathbf{S}^T \mathbf{A} \mathbf{S} \quad ; \quad \mathbf{S}^T \mathbf{S} = \mathbf{1}.$$

\mathbf{S} and \mathbf{H} are related by:

$$\mathbf{A} = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T = \mathbf{S} \mathbf{\Lambda}^{1/2} \mathbf{\Lambda}^{1/2} \mathbf{S}^T = \mathbf{H}^T \mathbf{H} \quad ; \quad \mathbf{H} = \mathbf{\Lambda}^{1/2} \mathbf{S}^T \quad ; \quad \mathbf{H}^{-1} = \mathbf{S} \mathbf{\Lambda}^{-1/2}$$

D.2. Normalization property

To verify the normalization property, we integrate the distribution:

$$\int_{\mathbb{R}^n} d\mathbf{X} f_{x_n}(\mathbf{X}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} d\mathbf{Y} \exp\left[-\frac{1}{2} \mathbf{Y}^T \mathbf{A} \mathbf{Y}\right] = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} d\mathbf{Y} \exp\left[-\frac{1}{2} \mathbf{Y}^T \mathbf{H}^T \mathbf{H} \mathbf{Y}\right]$$

Making the transformation $\mathbf{U} = \mathbf{H}\mathbf{Y}$, (with jacobian $\partial(\mathbf{Y}^T) / \partial(\mathbf{U}^T) = |\mathbf{H}^{-1}| = \frac{1}{|\mathbf{A}|^{1/2}}$)

we find

$$\int_{\mathbb{R}^n} d\mathbf{X} f_{x_n}(\mathbf{X}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} d\mathbf{U} \frac{1}{|\mathbf{A}|^{1/2}} \exp\left[-\frac{1}{2} \mathbf{U}^T \mathbf{U}\right] = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \frac{(2\pi)^{n/2}}{|\mathbf{A}|^{1/2}} = 1$$

D.3. Centered second order moment

$$\int_{\mathbb{R}^n} d\mathbf{X} (X_i - m_i)(X_j - m_j) f_{x_n}(\mathbf{X}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} d\mathbf{Y} Y_i Y_j \exp\left[-\frac{1}{2} \mathbf{Y}^T \mathbf{A} \mathbf{Y}\right]$$

Making the transformation $\mathbf{U} = \mathbf{H}\mathbf{Y}$, (with jacobian $\frac{\partial(\mathbf{Y}^T)}{\partial(\mathbf{U}^T)} = \frac{1}{|\mathbf{A}|^{1/2}}$) we find

$$\begin{aligned} \int_{\mathbb{R}^n} d\mathbf{X} (X_i - m_i)(X_j - m_j) f_{x_n}(\mathbf{X}) &= \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} d\mathbf{U} \frac{1}{|\mathbf{A}|^{1/2}} (\mathbf{H}^{-1}\mathbf{U})_i (\mathbf{H}^{-1}\mathbf{U})_j \exp\left[-\frac{1}{2} \mathbf{U}^T \mathbf{U}\right] \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} d\mathbf{U} \sum_k \mathbf{H}_{ik}^{-1} U_k \sum_l \mathbf{H}_{jl}^{-1} U_l \exp\left[-\frac{1}{2} \mathbf{U}^T \mathbf{U}\right] \\ &= \frac{1}{(2\pi)^{n/2}} \sum_k \sum_l \mathbf{H}_{ik}^{-1} \mathbf{H}_{jl}^{-1} \int_{\mathbb{R}^n} d\mathbf{U} U_k U_l \exp\left[-\frac{1}{2} \mathbf{U}^T \mathbf{U}\right] \\ &= \frac{1}{(2\pi)^{n/2}} \sum_k \sum_l \mathbf{H}_{ik}^{-1} \mathbf{H}_{jl}^{-1} \delta_{ij} (2\pi)^{n/2} \\ &= \sum_k \mathbf{H}_{ik}^{-1} \mathbf{H}_{jk}^{-1} = \sum_k (\mathbf{H}^{-1})_{ik} (\mathbf{H}^{-1})_{kj} = (\mathbf{H}^{-1} \mathbf{H}^{-1T})_{ij} = (\mathbf{A}^{-1})_{ij} = C_{ij}, \end{aligned}$$

so that, as expected,

$$\overline{(X_i - m_i)(X_j - m_j)} = C_{ij}.$$

D.4. Conditional distribution.

(Tong, Y. L., p35) Let \mathbf{X} be partitioned as $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$. If \mathbf{X} is multivariate normal with mean $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ and covariance matrix $\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}$, then the conditional distribution of \mathbf{X}_1 , given $\mathbf{X}_2 = \mathbf{x}_2$, is multivariate normal with mean $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{C}_{12}\mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ and covariance $\mathbf{C}_{11|2} = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$.

D.5. Marginal distribution

The marginal distribution of the subset $\mathbf{x}_1 = [x_1, \dots, x_{n_1}]$, of the multivariate normally distributed random vector $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]$, where $\mathbf{x}_2 = [x_{n_1+1}, \dots, x_{n_1+n_2}]$ and $n = n_1 + n_2$, is also multivariate normally distributed, with covariance matrix $\mathbf{C}_{11} = [E(Y_i Y_j)]_{n_1, n_1}$. The covariance matrix \mathbf{C}_{11} for \mathbf{x}_1 is expressed as a function of the covariance matrix \mathbf{C} for \mathbf{x} as follow:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^T & \mathbf{C}_{22} \end{bmatrix}$$

This property leads to the following equation

$$f_{\mathbf{x}_1}(\mathbf{X}_1) = \int_{\mathbf{R}^{n_2}} d\mathbf{X}_2 f_{\mathbf{x}}(\mathbf{X}) = \frac{|\mathbf{C}^{-1}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbf{R}^{n_2}} d\mathbf{X}_2 \exp\left[-\frac{1}{2}\mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y}\right]$$

$$= \frac{|\mathbf{C}_{11}^{-1}|^{1/2}}{(2\pi)^{n_1/2}} \exp\left[-\frac{1}{2} \mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1\right]$$

Proof:

Let $\mathbf{A} = \mathbf{C}^{-1}$, and define the following partition

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix}$$

We may write

$$\begin{aligned} \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y} &= \mathbf{Y}^T \mathbf{A} \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1^T & \mathbf{Y}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \\ &= \mathbf{Y}_1^T \mathbf{A}_{11} \mathbf{Y}_1 + 2 \mathbf{Y}_1^T \mathbf{A}_{12} \mathbf{Y}_2 + \mathbf{Y}_2^T \mathbf{A}_{22} \mathbf{Y}_2 \end{aligned}$$

Noting that \mathbf{A}_{22} is symmetric p.d. we can use a Cholesky factorization and write $\mathbf{Y}_2^T \mathbf{A}_{22} \mathbf{Y}_2 = \mathbf{Y}_2^T \mathbf{H}^T \mathbf{H} \mathbf{Y}_2 = \mathbf{U}^T \mathbf{U}$, where $\mathbf{U} = \mathbf{H} \mathbf{Y}_2$, $\mathbf{Y}_2 = \mathbf{H}^{-1} \mathbf{U}$ and \mathbf{H} is an upper triangular matrix. Using this relationship we have

$$\begin{aligned} \mathbf{Y}^T \mathbf{A} \mathbf{Y} &= \mathbf{Y}_1^T \mathbf{A}_{11} \mathbf{Y}_1 + 2 \underbrace{\mathbf{Y}_1^T \mathbf{A}_{12} \mathbf{H}^{-1}}_{\mathbf{V}^T} \mathbf{U} + \mathbf{U}^T \mathbf{U} \\ &= \mathbf{Y}_1^T \mathbf{A}_{11} \mathbf{Y}_1 - \mathbf{V}^T \mathbf{V} + \underbrace{(\mathbf{U} + \mathbf{V})^T}_{\mathbf{W}^T} \underbrace{(\mathbf{U} + \mathbf{V})}_{\mathbf{W}} \\ &= \mathbf{Y}_1^T \mathbf{A}_{11} \mathbf{Y}_1 - \mathbf{V}^T \mathbf{V} + \mathbf{W}^T \mathbf{W} \\ &= \mathbf{Y}_1^T \mathbf{A}_{11} \mathbf{Y}_1 - \mathbf{Y}_1^T \mathbf{A}_{12} \underbrace{\mathbf{H}^{-1} \mathbf{H}^{-1T}}_{\mathbf{A}_{22}^{-1}} \mathbf{A}_{21} \mathbf{Y}_1 + \mathbf{W}^T \mathbf{W} \\ &= \mathbf{Y}_1^T (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}) \mathbf{Y}_1 + \mathbf{W}^T \mathbf{W} \end{aligned}$$

Making the transformation $\mathbf{W} = \mathbf{U} + \mathbf{V} = \mathbf{H}\mathbf{Y}_2 + \mathbf{V}$, $\mathbf{Y}_2 = \mathbf{H}^{-1}(\mathbf{W} - \mathbf{V})$ (with Jacobian $\frac{\partial(\mathbf{Y}_2)}{\partial(\mathbf{W})} = \frac{1}{|\mathbf{H}|} = \frac{1}{|\mathbf{A}_{22}|^{1/2}}$) in the integral we get

$$\begin{aligned} \int_{\mathbb{R}^{n_2}} d\mathbf{X}_2 f_{\mathbf{X}}(\mathbf{X}) &= \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n_2}} d\mathbf{Y}_2 \exp\left[-\frac{1}{2}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}]\right] \\ &= \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n_2}} d\mathbf{W} \frac{1}{|\mathbf{A}_{22}|^{1/2}} \exp\left[-\frac{1}{2}[\mathbf{Y}_1^T (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}) \mathbf{Y}_1 + \mathbf{W}^T \mathbf{W}]\right] \\ &= \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \frac{1}{|\mathbf{A}_{22}|^{1/2}} \exp\left[-\frac{1}{2}[\mathbf{Y}_1^T (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}) \mathbf{Y}_1]\right] (2\pi)^{n_2/2} \\ &= \frac{1}{(2\pi)^{n_1/2}} \frac{|\mathbf{A}|^{1/2}}{|\mathbf{A}_{22}|^{1/2}} \exp\left[-\frac{1}{2}[\mathbf{Y}_1^T (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}) \mathbf{Y}_1]\right]. \end{aligned}$$

Now we note the following property of the inverse of a partitioned matrix:

$$\mathbf{A}^{-1} = \mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ -\mathbf{S}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{S}^{-1} \end{bmatrix},$$

where $\mathbf{S} = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}$ is the Schur complement. Using the properties of matrices and determinants we can see from the above equation that

$$\mathbf{C}_{22} = \mathbf{S}^{-1} = \left(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}\right)^{-1}, \text{ or equivalently}$$

$$\mathbf{C}_{22}^{-1} = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12},$$

and

$$|\mathbf{A}| = |\mathbf{A}_{11}| |\mathbf{S}| = \frac{|\mathbf{A}_{11}|}{|\mathbf{C}_{22}|}.$$

Similarly, considering $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ as a Schur complement we can write

$$\mathbf{C}_{11}^{-1} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21},$$

and

$$|\mathbf{A}| = \frac{|\mathbf{A}_{22}|}{|\mathbf{C}_{11}|}.$$

Substituting those relationships in the expression obtained for $\int_{\mathbf{R}^{n_2}} d\mathbf{X}_2 f_{\mathbf{x}}(\mathbf{X})$ we get the

required result that

$$f_{\mathbf{x}_1}(\mathbf{X}_1) = \int_{\mathbf{R}^{n_2}} d\mathbf{X}_2 f_{\mathbf{x}}(\mathbf{X}) = \frac{|\mathbf{C}_{11}^{-1}|^{1/2}}{(2\pi)^{n_1/2}} \exp\left[-\frac{1}{2}\mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1\right].$$

D.6. Approximation of the integral over a hyper rectangle of the subset of a random vector which is multivariate normal

PROPOSITION D.1: Again let $\mathbf{x} = [x_1, \dots, x_{n_1}, x_{n_1+1}, \dots, x_n]$ be a vector of random variables with multivariate normal distribution, and let $\mathbf{x}_1 = [x_1, \dots, x_{n_1}]$ and $\mathbf{x}_2 = [x_{n_1+1}, \dots, x_{n_1+n_2}]$, where $n = n_1 + n_2$ be subsets of the vector \mathbf{x} . Then the integral $\int_{\mathbf{I}} d\mathbf{X}_2 f_{\mathbf{x}}(\mathbf{X})$, where \mathbf{I} is a hyper rectangle in \mathbf{R}^{n_2} with lower left corner $\mathbf{a} = [a_1, \dots, a_{n_2}]$ and upper right corner $\mathbf{b} = [b_1, \dots, b_{n_2}]$, i.e. $\mathbf{I} = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_{n_2}, b_{n_2}]$, is given by

$$\int_{\mathbf{I}} d\mathbf{X}_2 f_{\mathbf{x}}(\mathbf{X}) = \frac{|\mathbf{C}^{-1}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbf{I}} d\mathbf{X}_2 \exp\left[-\frac{1}{2}\mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y}\right]$$

$$\approx \frac{1}{(2\pi)^{n_1/2} |\mathbf{C}_{11}|^{1/2}} \exp -\frac{1}{2} \mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1 \prod_{i=1}^{n_2} \frac{\text{erf}(\frac{\beta_i}{\sqrt{2}}) - \text{erf}(\frac{\alpha_i}{\sqrt{2}})}{2^{n_2}},$$

where $\mathbf{m} = E[\mathbf{x}]$, $\mathbf{Y} = \mathbf{X} - \mathbf{m}$, the covariance matrix is partitioned as follow

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^T & \mathbf{C}_{22} \end{bmatrix},$$

its inverse $\mathbf{A} = \mathbf{C}^{-1}$ is also partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix},$$

the coefficients a_i and b_i are elements of the vectors \mathbf{a} and \mathbf{b} given by

$$\boldsymbol{\alpha} = \mathbf{H}(\mathbf{a} - \mathbf{m}_2) + \mathbf{H}^{-1T} \mathbf{A}_{21} \mathbf{Y}_1, \text{ and}$$

$$\boldsymbol{\beta} = \mathbf{H}(\mathbf{b} - \mathbf{m}_2) + \mathbf{H}^{-1T} \mathbf{A}_{21} \mathbf{Y}_1,$$

and \mathbf{H} is an upper triangular matrix defined by $\mathbf{A}_{22} = \mathbf{H}^T \mathbf{H}$.

Derivation of the proposition: As shown above we may write

$$\mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y} = \mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1 + \mathbf{W}^T \mathbf{W},$$

where

$$\mathbf{W} = \mathbf{H} \mathbf{Y}_2 + \mathbf{H}^{-1T} \mathbf{A}_{21} \mathbf{Y}_1 = \mathbf{H}(\mathbf{X}_2 - \mathbf{m}_2) + \mathbf{H}^{-1T} \mathbf{A}_{21} \mathbf{Y}_1,$$

and \mathbf{H} is an upper triangular matrix defined by $\mathbf{A}_{22} = \mathbf{H}^T \mathbf{H}$

Making the transformation $\mathbf{W} = \mathbf{H}(\mathbf{X}_2 - \mathbf{m}_2) + \mathbf{H}^{-1T} \mathbf{A}_{21} \mathbf{Y}_1$,
 $\mathbf{X}_2 = \mathbf{m}_2 + \mathbf{H}^{-1}(\mathbf{W} - \mathbf{H}^{-1T} \mathbf{A}_{21} \mathbf{Y}_1)$ (with Jacobian $\frac{\partial(\mathbf{X}_2)}{\partial(\mathbf{W})} = \frac{1}{|\mathbf{H}|} = \frac{1}{|\mathbf{A}_{22}|^{1/2}}$) in the

integral we get

$$\begin{aligned} \int_{\mathbf{I}} d\mathbf{X}_2 f_{\mathbf{X}}(\mathbf{X}) &= \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbf{I}} d\mathbf{X}_2 \exp\left[-\frac{1}{2}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}]\right] \\ &\approx \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} \int_{I_{\mathbf{W}}} d\mathbf{W} \frac{1}{|\mathbf{A}_{22}|^{1/2}} \exp\left[-\frac{1}{2}[\mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1 + \mathbf{W}^T \mathbf{W}]\right] \\ &\approx \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2} |\mathbf{A}_{22}|^{1/2}} \exp\left[-\frac{1}{2} \mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1\right] \int_{I_{\mathbf{W}}} d\mathbf{W} \exp\left[-\frac{1}{2} \mathbf{W}^T \mathbf{W}\right], \end{aligned}$$

where the transformed domain is approximated by $I_{\mathbf{W}}$, the hyper-rectangle with lower left corner $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{n_2}]$ and upper right corner $\boldsymbol{\beta} = [\beta_1, \dots, \beta_{n_2}]$, defined by

$$\boldsymbol{\alpha} = \mathbf{H}(\mathbf{a} - \mathbf{m}_2) + \mathbf{H}^{-1T} \mathbf{M}_{21} \mathbf{Y}_1, \text{ and}$$

$$\boldsymbol{\beta} = \mathbf{H}(\mathbf{b} - \mathbf{m}_2) + \mathbf{H}^{-1T} \mathbf{M}_{21} \mathbf{Y}_1.$$

We now express the multiple integrals $\int_{I_{\mathbf{W}}} d\mathbf{W} \exp\left[-\frac{1}{2}[\mathbf{W}^T \mathbf{W}]\right]$ as the product of independent

integrals that can be evaluated using the error function. The error function is defined by

$$\frac{2}{\sqrt{\pi}} \int_0^x du \exp(-u^2) = \text{erf}(x).$$

It is convenient to note that we can also write

$$\int_a^b du \exp\left(-\frac{u^2}{2}\right) = \sqrt{\frac{\pi}{2}} \left[\operatorname{erf}\left(\frac{b}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a}{\sqrt{2}}\right) \right]$$

Using the above relationship we can write

$$\begin{aligned} \int_{I_{\mathbf{W}}} d\mathbf{W} \exp\left(-\frac{1}{2}[\mathbf{W}^T \mathbf{W}]\right) &= \prod_{i=1}^{n_2} \int_{\alpha_i}^{\beta_i} dW_i \exp\left(-\frac{1}{2}W_i^2\right) = \prod_{i=1}^{n_2} \sqrt{\frac{\pi}{2}} \left[\operatorname{erf}\left(\frac{\beta_i}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{\alpha_i}{\sqrt{2}}\right) \right] \\ &= (2\pi)^{\frac{n_2}{2}} \prod_{i=1}^{n_2} \frac{\operatorname{erf}\left(\frac{\beta_i}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{\alpha_i}{\sqrt{2}}\right)}{2^{n_2}}. \end{aligned}$$

Therefore we get the required approximation:

$$\int_{\mathbf{I}} d\mathbf{X}_2 f_{\mathbf{x}}(\mathbf{X}) \approx \frac{1}{(2\pi)^{n_1/2} |\mathbf{C}_{11}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{Y}_1^T \mathbf{C}_{11}^{-1} \mathbf{Y}_1\right) \prod_{i=1}^{n_2} \frac{\operatorname{erf}\left(\frac{\beta_i}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{\alpha_i}{\sqrt{2}}\right)}{2^{n_2}}.$$

D.7. Usefull integrals

$$\frac{2}{\sqrt{\pi}} \int_0^x du \exp(-u^2) = \operatorname{erf}(x),$$

$$\int_a^b du \exp\left(-\frac{u^2}{2}\right) = \sqrt{\frac{\pi}{2}} \left[\operatorname{erf}\left(\frac{b}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a}{\sqrt{2}}\right) \right],$$

$$\int_{-\infty}^{\infty} du \exp(-u^2) = \sqrt{\pi},$$

$$\int_{-\infty}^{\infty} du \exp\left(-\frac{u^2}{2}\right) = \sqrt{2\pi},$$

APPENDIX E: THE THEORY OF SPACE TRANSFORMATIONS AND THEIR NUMERICAL IMPLEMENTATION

E.1. A Brief Review of the Theory of Space Transformations

In the analysis of subsurface processes, field parameters such as permeability and porosity are commonly treated as spatial random fields (e.g., Neuman, 1982; Naff and Vecchia, 1986), the realizations of which are multidimensional functions from a statistical ensemble that appear with frequency determined by the multivariate probability distribution. Thus, it is possible to obtain estimates that account for the variability and the uncertainties inherent in most environmental processes. In general, ST are mathematical operations that map a process defined on a multidimensional space onto a space of reduced dimensionality. The ST operators can be applied to random field realizations, if the latter are integrable functions, or to the statistical moments of the random field (Christakos, 1987; 1992). Below we briefly review the definitions of ST operators and some of their properties.

Let \mathbf{R}^n be the n -dimensional Euclidean space, $\mathbf{s} = (s_1, s_2, \dots, s_n) \in \mathbf{R}^n$ denote the position vector, and $\boldsymbol{\theta}$ denote a direction vector on the generalized unit sphere S^{n-1} . Let $f_n(\mathbf{s})$ be a function defined on \mathbf{R}^n that is either rapidly decreasing or has a compact support. Then one can define a mapping T_n^1 from the space \mathbf{R}^n into the space $\mathbf{R}^1 \times S^{n-1}$ as follows

$$T_n^1: f_n(\mathbf{s}) \rightarrow \int \mathbf{d}\mathbf{s} f_n(\mathbf{s}) \delta(\mathbf{p} - \mathbf{s} \cdot \boldsymbol{\theta}), \quad (\text{E.1})$$

where $\mathbf{p} \in \mathbf{R}^1$, and $\delta(\cdot)$ is the singular distribution known as the Dirac delta function. The image of the mapping for all the direction vectors on S^{n-1} and all real numbers \mathbf{p} is known as the Radon transform (Helgason, 1980; Deans, 1993) or as the plane wave

integral (John, 1955). The T_n^1 transformation in a specific direction represents the projection of the n -dimensional function on the hyperplanes that are perpendicular to this direction. The mapping (E.1) is an ST expressed as the integral of the function $f_n(\mathbf{s})$ over the hyperplane that is perpendicular to the direction vector $\boldsymbol{\theta}$ and is located at a distance p from the origin (this ST is usually called an ST-1 in order to be distinguished from other kinds of ST; see, also, below).

The infinite line in the direction of the vector $\boldsymbol{\theta}$ is called an ST line. The operator T_n^1 reduces an n -dimensional function to an one-dimensional function defined as

$$T_n^1[f_n](p, \boldsymbol{\theta}) = \hat{f}_{1, \boldsymbol{\theta}}(p). \quad (\text{E.2})$$

In the case of an isotropic function $f_n(\mathbf{s})$ with infinite support, the T_n^1 is also isotropic, i.e., the function $\hat{f}_{1, \boldsymbol{\theta}}(p)$ is independent of the orientation of the direction vector $\boldsymbol{\theta}$. This is not true in the case of a finite support that lacks spherical symmetry. Fig. E.1 shows the intersection of a transformation plane with a compact rectangular support. The equations that determine the intersections of the transformation planes in a cubic support are derived (Hristopulos et al.; 1999). The ST $\hat{f}_{1, \boldsymbol{\theta}}(p)$ is equal to the integral of $f_3(\mathbf{s})$ over the planar domain defined by the intersection of the support with the infinite plane that is perpendicular to $\boldsymbol{\theta}$ and is located at distance p from the origin. A different direction vector will, in general, lead to a different ST function. Note that the finite support acts as a spatial filter for the n -dimensional function, that permits the ST integral operator (E.1) to be always well-defined. In exchange, the compact support introduces boundary effects in the space-transformed functions.

We present without proof certain properties for the T_n^1 that will be used later; proofs can be obtained directly from the T_n^1 -definition using the properties of generalized

functions (an excellent review of the theory of generalized functions may be found in Gel'fand and Shilov, 1964):

(a) scaling

$$T_n^1[f_n(\lambda \mathbf{s})] = \frac{1}{\lambda^{n-1}} \hat{f}_{1,\theta}(\lambda \mathbf{p}); \quad (\text{E.3})$$

(b) shifting

$$T_n^1[f_n(\mathbf{s} - \mathbf{a})] = \hat{f}_{1,\theta}(\mathbf{p} - \theta \cdot \mathbf{a}); \quad (\text{E.4})$$

and

(c) complementarity

$$\hat{f}_{1,\theta}(\mathbf{p}) = \hat{f}_{1,-\theta}(-\mathbf{p}). \quad (\text{E.5})$$

The complementarity property reflects the fact that the mapping from the space $\mathbf{R} \times \mathbf{S}^{n-1}$ onto the space of the n -hyperplanes leads to a double covering of the hyperplane space. Single covering is obtained by restricting the transformation to positive projections $\mathbf{p} > 0$ or to one hemisphere of the unit sphere.

Another ST is the Ψ_n^1 operator defined by

$$\Psi_n^1[f_n](\mathbf{p}, \theta) \equiv \Omega T_n^1[f_n](\mathbf{p}, \theta) = f_{1,\theta}(\mathbf{p}); \quad (\text{E.6})$$

the Ω is a differential operator such that

$$\Omega = \frac{(-1)^m}{2(2\pi)^{2m}} \times \begin{cases} S_{2m+1} \frac{\partial^{2m}}{\partial p^{2m}} [\cdot] & n = 2m + 1 \\ 2\pi S_{2m} \mathbf{H}\left\{\frac{\partial^{2m-1}}{\partial p^{2m-1}}\right\}[\cdot] & n = 2m \end{cases}, \quad (\text{E.7})$$

where $S_n = (2\pi)^{n/2} / \Gamma(\frac{n}{2})$ represents the surface area of the n -dimensional unit sphere S^{n-1} , Γ is the gamma function and \mathbf{H} denotes the Hilbert transform (the Ψ_n^1 is sometimes called the ST-2).

The inverse ST Ψ_1^n and T_1^n of the ST Ψ_n^1 and T_n^1 , respectively, exist and are defined by

$$\Psi_1^n [f_{1,\theta}](\mathbf{s}) = \frac{1}{S_n} \int_{S_n} f_{1,\theta}(\mathbf{s} \cdot \boldsymbol{\theta}) \mathbf{d}\boldsymbol{\theta} = f_n(\mathbf{s}) \quad (\text{E.8})$$

and

$$T_1^n [\hat{f}_{1,\theta}](\mathbf{s}) = \Psi_1^n \Omega[\hat{f}_{1,\theta}](\mathbf{s}) = f_n(\mathbf{s}). \quad (\text{E.9})$$

In the spectral domain, the ST and their inverses turn out to have simple algebraic forms. In particular, for the T_n^1 we have

$$T_n^1 [\tilde{f}_n](\boldsymbol{\omega}, \boldsymbol{\theta}) = \tilde{f}_n(\mathbf{w}) = \hat{\tilde{f}}_{1,\theta}(\boldsymbol{\omega}) \quad (\text{E.10})$$

along the wave vector $\mathbf{w} = \boldsymbol{\omega}\boldsymbol{\theta}$, where \tilde{f} denotes the Fourier transform of f . For the inverse ST T_1^n we have

$$\mathbf{T}_1^n[\hat{\mathbf{f}}_{1,\theta}](\mathbf{w}) = \hat{\mathbf{f}}_{1,\theta}(\boldsymbol{\omega}) = \tilde{\mathbf{f}}_n(\mathbf{w}). \quad (\text{E.11})$$

Similarly, for the Ψ_n^1 it can be shown that the following spectral relation holds

$$\Psi_n^1[\tilde{\mathbf{f}}_n](\boldsymbol{\omega}) = \frac{\mathbf{S}_n |\boldsymbol{\omega}|^{n-1} \tilde{\mathbf{f}}_n(\mathbf{w})}{(2\pi)^{n-1}} = \hat{\mathbf{f}}_{1,\theta}(\boldsymbol{\omega}), \quad (\text{E.12})$$

while for the Ψ_1^n the corresponding relation is

$$\Psi_1^n[\hat{\mathbf{f}}_{1,\theta}](\mathbf{w}) = \frac{(2\pi)^{n-1} \hat{\mathbf{f}}_{1,\theta}(\boldsymbol{\omega})}{\mathbf{S}_n |\boldsymbol{\omega}|^{n-1}} = \tilde{\mathbf{f}}_n(\mathbf{w}). \quad (\text{E.13})$$

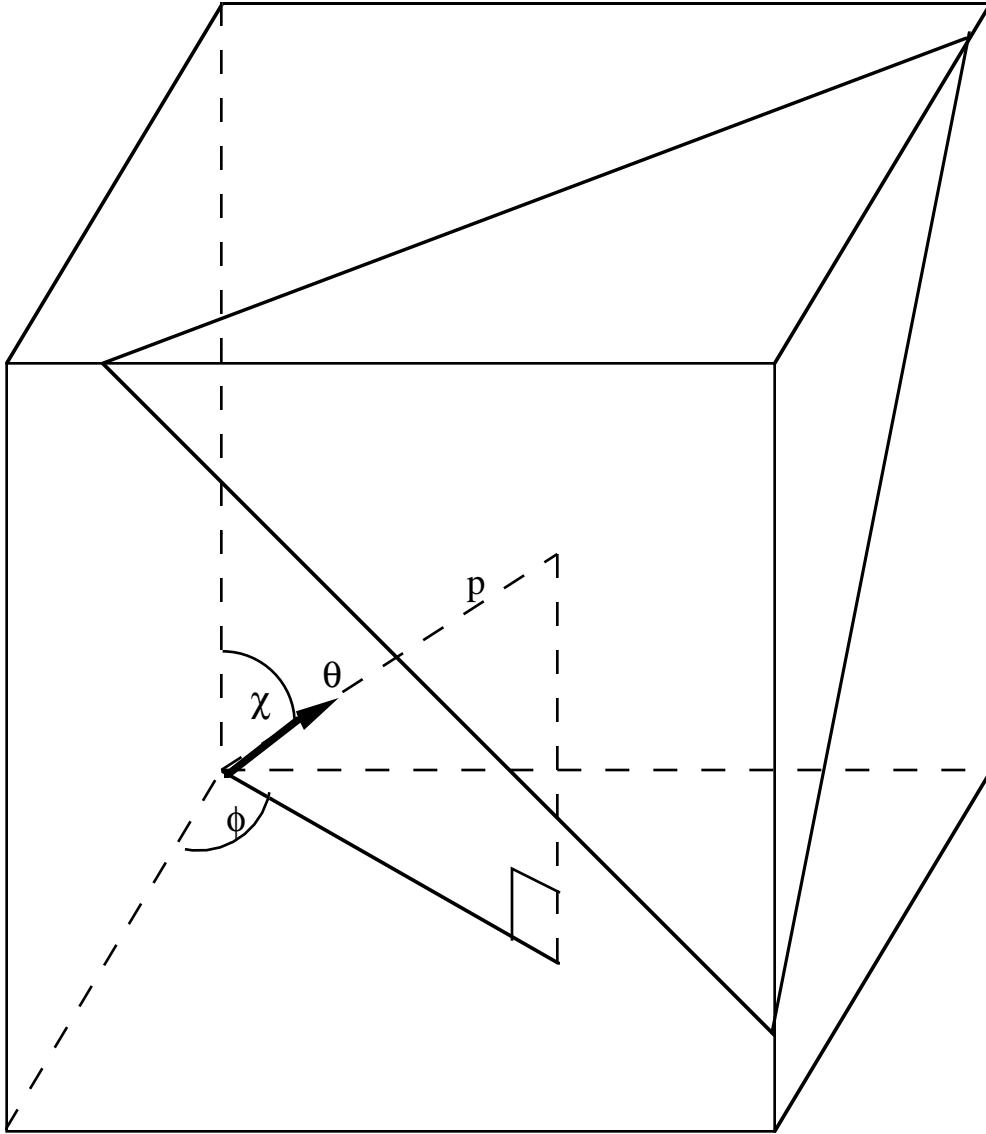


Figure E.1: A cubic three-dimensional domain and the T_3^1 transformation plane in the direction θ with projection length p . The $T_3^1[\cdot](p, \theta)$ represents the integral of a three-dimensional function over the plane.

APPENDIX F: KNOWLEDGE PROCESSING RULE FOR SOFT INTERVAL DATA

Let's define the random variable x taking values x and the random vectors \mathbf{y} and \mathbf{z} taking values $\boldsymbol{\Psi}$ and $\boldsymbol{\zeta}$. The cumulative distribution function (cdf) of x , \mathbf{y} and \mathbf{z} , $F_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta})$, is defined as the following probability

$$F_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta}) = P(x < \chi, \mathbf{y} < \boldsymbol{\Psi}, \mathbf{z} < \boldsymbol{\zeta}).$$

When it exists, the probability distribution function (pdf) $f_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta})$ of the random variables x , \mathbf{y} and \mathbf{z} is defined as

$$\begin{aligned} f_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta}) &= \frac{\partial \dots \partial F_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta})}{\partial \chi \partial \boldsymbol{\Psi} \partial \boldsymbol{\zeta}} \\ &= \lim_{d\chi, d\boldsymbol{\Psi}, d\boldsymbol{\zeta} \rightarrow 0} \frac{F_{xyz}(\chi + d\chi, \boldsymbol{\Psi} + d\boldsymbol{\Psi}, \boldsymbol{\zeta} + d\boldsymbol{\zeta}) - F_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta})}{d\chi d\boldsymbol{\Psi} d\boldsymbol{\zeta}} \\ &= \lim_{d\chi, d\boldsymbol{\Psi}, d\boldsymbol{\zeta} \rightarrow 0} \frac{1}{d\chi d\boldsymbol{\Psi} d\boldsymbol{\zeta}} P(\chi < x < \chi + d\chi, \boldsymbol{\Psi} < \mathbf{y} < \boldsymbol{\Psi} + d\boldsymbol{\Psi}, \boldsymbol{\zeta} < \mathbf{z} < \boldsymbol{\zeta} + d\boldsymbol{\zeta}). \end{aligned}$$

Similarly it is natural to write

$$f_{yz}(\boldsymbol{\Psi}, \boldsymbol{\zeta}) = \lim_{d\boldsymbol{\Psi}, d\boldsymbol{\zeta} \rightarrow 0} \frac{1}{d\boldsymbol{\Psi} d\boldsymbol{\zeta}} P(\boldsymbol{\Psi} < \mathbf{y} < \boldsymbol{\Psi} + d\boldsymbol{\Psi}, \boldsymbol{\zeta} < \mathbf{z} < \boldsymbol{\zeta} + d\boldsymbol{\zeta}).$$

Now that the pdfs $f_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta})$ and $f_{yz}(\boldsymbol{\Psi}, \boldsymbol{\zeta})$ have been expressed in terms of limits of probability, let's express the conditional pdf $f(\chi | \boldsymbol{\Psi}, \boldsymbol{\zeta}) = \frac{f_{xyz}(\chi, \boldsymbol{\Psi}, \boldsymbol{\zeta})}{f_{yz}(\boldsymbol{\Psi}, \boldsymbol{\zeta})}$ in terms of probability. Using conditional probability we can write

$$\begin{aligned} &P(\chi < x < \chi + d\chi \mid \boldsymbol{\Psi} < \mathbf{y} < \boldsymbol{\Psi} + d\boldsymbol{\Psi}, \boldsymbol{\zeta} < \mathbf{z} < \boldsymbol{\zeta} + d\boldsymbol{\zeta}) \\ &= \frac{P(\chi < x < \chi + d\chi, \boldsymbol{\Psi} < \mathbf{y} < \boldsymbol{\Psi} + d\boldsymbol{\Psi}, \boldsymbol{\zeta} < \mathbf{z} < \boldsymbol{\zeta} + d\boldsymbol{\zeta})}{P(\boldsymbol{\Psi} < \mathbf{y} < \boldsymbol{\Psi} + d\boldsymbol{\Psi}, \boldsymbol{\zeta} < \mathbf{z} < \boldsymbol{\zeta} + d\boldsymbol{\zeta})}. \end{aligned}$$

By multiplying left and right by $1/d\chi$ and taking the limit when $d\chi$, $d\psi$ and $d\zeta$ tend to 0 we have

$$\begin{aligned}
& \lim_{d\chi, d\psi, d\zeta \rightarrow 0} \frac{1}{d\chi} P(\chi < x < \chi + d\chi \mid \psi < y < \psi + d\psi, \zeta < z < \zeta + d\zeta) \\
&= \frac{\lim_{d\chi, d\psi, d\zeta \rightarrow 0} \frac{1}{d\chi d\psi d\zeta} P(\chi < x < \chi + d\chi, \psi < y < \psi + d\psi, \zeta < z < \zeta + d\zeta)}{\lim_{d\psi, d\zeta \rightarrow 0} \frac{1}{d\psi d\zeta} P(\psi < y < \psi + d\psi, \zeta < z < \zeta + d\zeta)} \\
&= \frac{f_{xyz}(\chi, \psi, \zeta)}{f_{yz}(\psi, \zeta)} \\
&= f(\chi \mid \psi, \zeta).
\end{aligned}$$

The above result simply states that the pdf of x given the knowledge that $y = \psi$ and $z = \zeta$ can be expressed in terms of a limit of probability as follow

$$f(\chi \mid \psi, \zeta) = \lim_{d\chi, d\psi, d\zeta \rightarrow 0} \frac{1}{d\chi} P(\chi < x < \chi + d\chi \mid \psi < y < \psi + d\psi, \zeta < z < \zeta + d\zeta).$$

Note that we take limit of a probability normalized by $d\chi$ only. Hence, we can generalize the above definition to the case of the pdf of x given the knowledge that $y = \psi$ and $\zeta_1 \leq z \leq \zeta_u$ as follow

$$\begin{aligned}
f(\chi \mid \psi, \zeta_1 \leq z \leq \zeta_u) &= \lim_{d\chi, d\psi \rightarrow 0} \frac{1}{d\chi} P(\chi < x < \chi + d\chi \mid \psi < y < \psi + d\psi, \zeta_1 \leq z \leq \zeta_u) \\
&= \frac{\lim_{d\chi, d\psi \rightarrow 0} \frac{1}{d\chi d\psi} P(\chi < x < \chi + d\chi, \psi < y < \psi + d\psi, \zeta_1 \leq z \leq \zeta_u)}{\lim_{d\psi \rightarrow 0} \frac{1}{d\psi} P(\psi < y < \psi + d\psi, \zeta_1 \leq z \leq \zeta_u)},
\end{aligned}$$

which can at once be written in terms of $f_{xyz}(\chi, \psi, \zeta)$ and $f_{yz}(\psi, \zeta)$ as

$$f(\chi \mid \psi, \zeta_1 \leq z \leq \zeta_u) = \frac{\int_{\zeta_1}^{\zeta_u} d\zeta f_{xyz}(\chi, \psi, \zeta)}{\int_{\zeta_1}^{\zeta_u} d\zeta f_{yz}(\psi, \zeta)}.$$

APPENDIX G: PROOF OF CONFIDENCE SETS

In this Appendix we briefly discuss the derivation of Eq. (xx). Consider the high probability set $\Lambda_\beta = \{\boldsymbol{x}_k: f_K(\boldsymbol{x}_k)/f_K(\hat{\boldsymbol{x}}_k) \geq 1 - \beta\}$ and any other set Φ_η' such that $P(\Phi_\eta') = P(\Lambda_\beta) = \eta$. The sets Λ_β and Φ_η' can be written as the union of two disjoint sets, i.e., $\Lambda_\beta = (\Lambda_\beta \cap \Phi_\eta') \cup (\Lambda_\beta - \Phi_\eta')$ and $\Phi_\eta' = (\Lambda_\beta \cap \Phi_\eta') \cup (\Phi_\eta' - \Lambda_\beta)$. Since $P(\Lambda_\beta) = P(\Phi_\eta')$, it follows that $P(\Lambda_\beta - \Phi_\eta') = P(\Phi_\eta' - \Lambda_\beta)$. But $(\Lambda_\beta - \Phi_\eta') \subset \Lambda_\beta$ and $(\Phi_\eta' - \Lambda_\beta) \subset \Lambda_\beta^c$, so that from the definition of high probability density set we have

$$\left. \begin{aligned} \forall \boldsymbol{x}_k \in (\Lambda_\beta - \Phi_\eta'), \quad f_K(\boldsymbol{x}_k) &\geq (1 - \beta)f_K(\hat{\boldsymbol{x}}_k) \\ \forall \boldsymbol{x}_k \in (\Phi_\eta' - \Lambda_\beta), \quad f_K(\boldsymbol{x}_k) &\leq (1 - \beta)f_K(\hat{\boldsymbol{x}}_k) \end{aligned} \right\} \quad (\text{G.1})$$

Using this result we can write

$$\left. \begin{aligned} P(\Lambda_\beta - \Phi_\eta') &= \int_{\boldsymbol{x}_k \in (\Lambda_\beta - \Phi_\eta')} d\boldsymbol{x}_k f_K(\boldsymbol{x}_k) \geq (1 - \beta)f_K(\hat{\boldsymbol{x}}_k) \|\Lambda_\beta - \Phi_\eta'\| \\ P(\Phi_\eta' - \Lambda_\beta) &= \int_{\boldsymbol{x}_k \in (\Phi_\eta' - \Lambda_\beta)} d\boldsymbol{x}_k f_K(\boldsymbol{x}_k) \leq (1 - \beta)f_K(\hat{\boldsymbol{x}}_k) \|\Phi_\eta' - \Lambda_\beta\| \end{aligned} \right\} \quad (\text{G.2})$$

But since $P(\Phi_\eta' - \Lambda_\beta) = P(\Lambda_\beta - \Phi_\eta')$, we get $\|\Lambda_\beta - \Phi_\eta'\| \leq \|\Phi_\eta' - \Lambda_\beta\|$. This implies that $\|\Lambda_\beta\| \leq \|\Phi_\eta'\|$ for any set Φ_η' such that $P(\Phi_\eta') = P(\Lambda_\beta) = \eta$, which completes the proof that Λ_β is the confidence set Φ_η with confidence level $\eta = P(\Lambda_\beta)$.