Concomitant of Ordered Multivariate Normal Distribution with Application to Parametric Inference

Eric Chaumette and François Vincent

Université de Toulouse/ISAE-Supaéro

Abstract

In statistics, the concept of a concomitant, also called the induced order statistic, arises when one sorts the members of a random sample according to corresponding values of another random sample. Indeed, multivariate order statistics induced by the ordering of linear combinations of the components arises naturally in many instances. As a contribution, we provide a general second-order statistical prediction of concomitant of order statistics for multivariate normal distribution, generalizing earlier works. We exemplify its usefulness in parametric inference via an example related to deterministic estimation.

On ordered maximum likelihood estimates (MLEs): case of a single parameter

- Let us consider the observation model formed from a linear superposition of M individual signals and noise:

$$\boldsymbol{y}(l) = \boldsymbol{H}(\theta)\,\boldsymbol{x}(l) + \boldsymbol{v}(l),\ 1 \le l \le L,\ \boldsymbol{y}(l) \in \mathbb{C}^N,\ \boldsymbol{x}_l \in \mathbb{C}^M, \quad (1)$$

$\boldsymbol{H}(\theta) = [\boldsymbol{h}(\theta_1), \dots, \boldsymbol{h}(\theta_M)]$ and $\boldsymbol{h}(\ )$ is a vector of N parametric functions depending on a parameter θ, $\boldsymbol{v}(l)$ are i.i.d. Gaussian complex circular noises.

- Since (1) is invariant over permutation of signal sources amplitude $\boldsymbol{x}(l)$, i.e. for any permutation matrix $\boldsymbol{P}_i \in \mathbb{R}^{M\times M}$:

$$\boldsymbol{y}(l) = (\boldsymbol{H}(\theta)\,\boldsymbol{P}_i)(\boldsymbol{P}_i\boldsymbol{x}(l)) + \boldsymbol{v}(l),$$

it is well known that (1) is an ill-posed unidentifiable estimation problem.

- **Definition:** *the ordered values of a sample of observations are called the order statistics of the sample: if* $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_M)^T$ *is a random vector, then* $\boldsymbol{\theta}_{(M)} = (\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(M)})^T$ *denotes the vector of order statistics induced by* $\boldsymbol{\theta}$ *where* $\theta_{(1)} \le \theta_{(2)} \le \dots \le \theta_{(M)}$.
- (1) can be regularized by imposing the ordering of the unknown parameters θ_m: $\boldsymbol{\theta} \triangleq \boldsymbol{\theta}_{(M)}$.Therefore in the MSE sense, the correct statistical prediction is given by $E\left[\left(\widehat{\theta}_{(m)} - \theta_m\right)^2\right], 1 \le m \le M$.
- Under reasonably general conditions on the observation model, MLEs are asymptotically Gaussian distributed when the number of independent observation tends to infinity.
- Nevertheless a close look at the derivations of these results reveals an implicit hypothesis: the asymptotic condition of operation considered yields resolvable estimates, what prevents from estimates re-ordering. Therefore, under this implicit hypothesis $\widehat{\boldsymbol{\theta}}_{(M)} = \widehat{\boldsymbol{\theta}}$.
- However when the condition of operation degrades, distribution spread and/or location bias of each $\widehat{\theta}_m$ increase and the hypothesis of resolvable estimates does not hold any longer yielding observation samples for which $\widehat{\boldsymbol{\theta}}_{(M)} \neq \widehat{\boldsymbol{\theta}}$.

Asymptotic conditional model with 2 sources: $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{CRB}_{\boldsymbol{\theta}})$

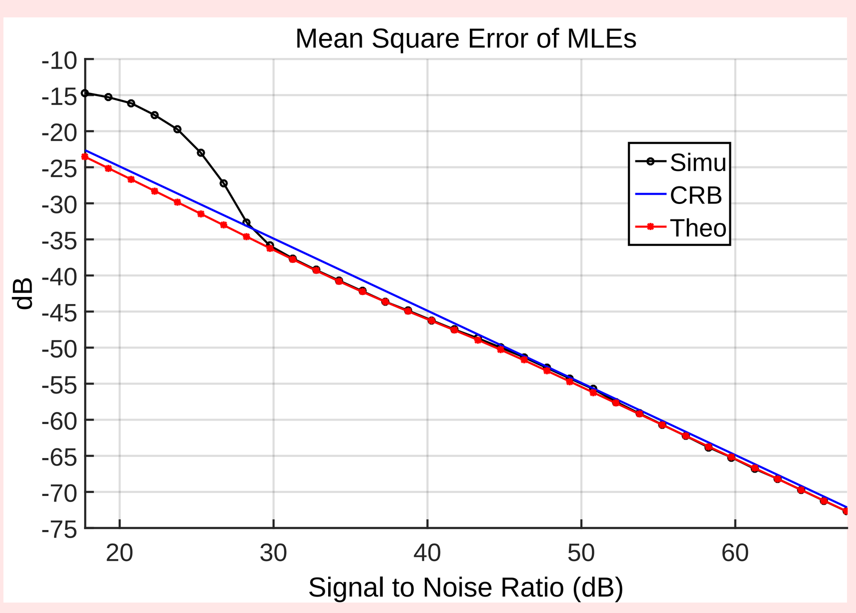
- If $\boldsymbol{CRB}_{\boldsymbol{\theta}} = \sigma^2\left((1-\rho)\,\boldsymbol{I}_2 + \rho\boldsymbol{1}_2\boldsymbol{1}_2^T\right)$, then $\sigma_{d\widehat{\theta}} = \sqrt{2\sigma^2(1-\rho)}$ and:

$$MSE\left[\widehat{\theta}_{(m)}\right] \triangleq E\left[\left(\widehat{\theta}_{(m)} - \theta_m\right)^2\right] = Var\left[\widehat{\theta}_m\right] - \sigma^2_{d\widehat{\theta}}\tau h(\tau), \quad (2a)$$

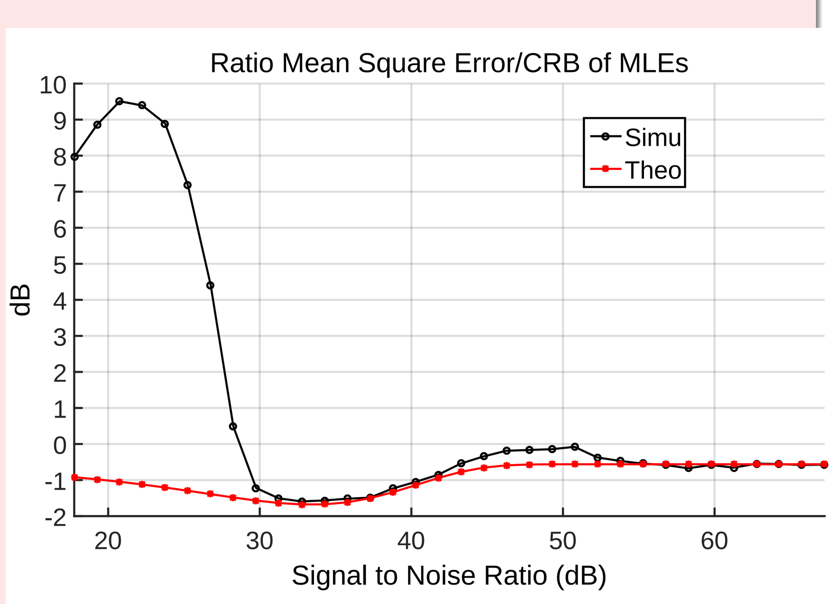
where $\tau = (\theta_2 - \theta_1)/\sigma_{d\widehat{\theta}}$, $h(y) = E\left[v1_{\{v\ge y\}}\right] - y\mathcal{P}(v \ge y)$, $v \sim \mathcal{N}(0,1)$. An interesting feature is the MSE shrinkage factor:

$$MSE\left[\widehat{\theta}_{(m)}\right]/Var\left[\widehat{\theta}_m\right] = 1 - 2(1-\rho)\,\tau h(\tau). \quad (2b)$$

- High resolution scenario: 2 tones of equal power with opposite frequencies where the observation model (1) is deterministic: $\boldsymbol{a}(\theta)^T = \left(1, e^{j2\pi\theta}, \dots, e^{j2\pi(N-1)\theta}\right)$, $N = 8$, $L = 2$, $\theta_2 - \theta_1 = \frac{1}{12N}$, $\boldsymbol{C}_{\boldsymbol{v}} = \boldsymbol{I}_2$, $\boldsymbol{C}_{\boldsymbol{x}} = \frac{SNR}{N}\left(\left(1+\frac{1}{8}\right)\boldsymbol{I}_2 - \frac{1}{8}\boldsymbol{1}_2\boldsymbol{1}_2^T\right)$ where SNR is measured at the output of the frequency matched filter.



Average MSE (2a)



Average MSE shrinkage factor (2b)

On ordered MLEs: case of a vector of parameters

- If $\boldsymbol{H}(\boldsymbol{\theta}) \to \boldsymbol{H}(\boldsymbol{\Theta}) = [\boldsymbol{h}(\boldsymbol{\theta}_1)\ \dots\ \boldsymbol{h}(\boldsymbol{\theta}_M)]$ where $\boldsymbol{h}(\ )$ is a vector of N parametric functions depending on a vector of P unknown parameters $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^P$, $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1\ \dots\ \boldsymbol{\theta}_M]$, (1) is still invariant over permutation of signal sources amplitude $\boldsymbol{x}(l)$.
- (1) is regularized by imposing the ordering of the unknown parameters $\{\boldsymbol{\theta}_m\}_{m=1}^M$.
- A natural ordering of $\{\boldsymbol{\theta}_m\}_{m=1}^M$ arises in the computation of MLEs. For instance, if (1) is a Gaussian conditional model:

$$\widehat{\boldsymbol{\Theta}} = \left[\widehat{\boldsymbol{\theta}}_1\ \dots\ \widehat{\boldsymbol{\theta}}_M\right] = \arg\max_{\boldsymbol{\Theta}}\left\{\boldsymbol{\Pi}_{\boldsymbol{H}(\boldsymbol{\Theta})}\widehat{\boldsymbol{R}}_{\boldsymbol{y}}\right\},\ \boldsymbol{\Pi}_{\boldsymbol{A}} = \boldsymbol{A}\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H, \quad (3a)$$

which can be solved only by numerical search techniques.

- A commonly used search technique is the conversion of a PM-dimensionnal search grid over Ω^M into a 1-dimensionnal search grid. For example, if $P = 2$ and $\boldsymbol{\theta} \in \Omega = [a_1, b_1] \times [a_2, b_2]$, a rectangular search grid over Ω is:

$$\mathcal{G} = \left\{\begin{pmatrix} a_1 + i_1\delta_1 \\ a_2 + i_2\delta_2 \end{pmatrix}, \middle|\ \begin{matrix} \delta_1 = (b_1 - a_1)/I_1,\ 0 \le i_1 \le I_1 \\ \delta_2 = (b_2 - a_2)/I_2,\ 0 \le i_2 \le I_2 \end{matrix}\right\} \quad (3b)$$

and convert each $\boldsymbol{\theta} \in \mathcal{G}$ into a linear search index $s = i_1 + (I_1 + 1)\,i_2$.

- Therefore, in practice (3a) becomes:

$$\widehat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}\left(\widehat{\boldsymbol{s}}_{(M)}\right),\ \widehat{\boldsymbol{s}} = \arg\max_{\boldsymbol{s}}\left\{\boldsymbol{\Pi}_{\boldsymbol{H}(\boldsymbol{\Theta}(\boldsymbol{s}))}\widehat{\boldsymbol{R}}_{\boldsymbol{y}}\right\}, \quad (3c)$$

which, as well, solves the issue of model identifiability.

On concomitant and ordered MLEs

- **Definition:** *Let us consider M random vectors with P components: $\{\boldsymbol{\theta}_m\}_{m=1}^M$. Let $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1\ \dots\ \boldsymbol{\theta}_M]$. If:*

$$\boldsymbol{s} = \left(\boldsymbol{\theta}_1^T\boldsymbol{a}, \dots, \boldsymbol{\theta}_M^T\boldsymbol{a}\right)^T = \boldsymbol{\Theta}^T\boldsymbol{a},\ \boldsymbol{a} \in \mathcal{M}_{\mathbb{R}}(P,1), \quad (4a)$$

then, the concomitants of $\boldsymbol{s}_{(M)} = \left(s_{(1)}, \dots, s_{(M)}\right)$ *are defined as:*

$$\boldsymbol{\Theta}_{[M]} = \left[\boldsymbol{\theta}_{[1]}\ \dots\ \boldsymbol{\theta}_{[M]}\right]\ \mid \boldsymbol{\theta}_{[m]} = \boldsymbol{\theta}_{m'} \Leftrightarrow s_{(m)} = s_{m'}. \quad (4b)$$

medskip

- If δ_1 and δ_2 are small enough, then:

$$s \simeq \boldsymbol{\theta}^T\boldsymbol{a} - s_0,\ \boldsymbol{a}^T = (1/\delta_1, (I_1+1)/\delta_2),\ s_0 = a_1/\delta_1 + (I_1+1)\,a_2/\delta_2. \quad (5)$$

Since the ordering does not depend on s_0, $\widehat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}\left(\widehat{\boldsymbol{s}}_{(M)}\right)$ are induced order statistic of $\widehat{\boldsymbol{s}}_{(M)}$, that is concomitants of $\widehat{\boldsymbol{s}}_{(M)}$.

Asymptotic conditional model with 2 sources (cont.)

- Let us consider a radar system consisting of a 1-element antenna array receiving scaled, timedelayed, and Doppler-shifted echoes of a known complex bandpass signal $e(t)\,e^{-j2\pi f_c t}$, where f_c is the carrier frequency.
- A standard observation model of a radar antenna receiving a pulse train of I pulses of duration δt_0 and bandwidth B, with a pulse repetition interval δt is given by (1) where $L = 1$, $N = \lfloor \delta t/B \rfloor$, $\boldsymbol{\theta}^T = (\tau, \omega)$, $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{\psi}(\omega) \otimes \boldsymbol{\phi}(\tau)$, $\boldsymbol{\psi}(\omega)^T = \left(1, \dots, e^{j2\pi\omega(I-1)\delta t}\right)$, $\boldsymbol{\phi}(\tau)^T = \left(e(-\tau), \dots, e\left(\frac{N-1}{B} - \tau\right)\right)$, τ and ω denoting the delay and the Doppler-shift associated to a target.
- The MLEs of $\boldsymbol{\Theta}$ are asymptotically efficient and Gaussian, and for 2 targets:

$$\boldsymbol{C}_{\boldsymbol{v}_{\widehat{\Theta}}} = \boldsymbol{CRB}_{\boldsymbol{v}_{\Theta}} = 2Re\left\{\boldsymbol{J}(\boldsymbol{\Theta}) \odot \left(\left(\boldsymbol{x}_1^T\boldsymbol{x}_1^*\right) \otimes \boldsymbol{1}_{2\times 2}\right)\right\}^{-1},\ \boldsymbol{v}_{\boldsymbol{\Theta}} = \boldsymbol{vec}(\boldsymbol{\Theta}),$$

where $\boldsymbol{J}(\boldsymbol{\Theta})$ is given in [Menni et al].

- We consider a high resolution scenario in terms of $\boldsymbol{\theta}$, that is a small Doppler-Shift $d\omega = 1/(12I)$ $(I = 8)$ and a small delays difference $d\tau = 1/(8B)$ $(\delta t_0 = 32/B)$. $e(t)$ is a linear chirp.
- The empirical MSE are assessed with 10^5 Monte-Carlo trials from the normally distributed vector associated with the asymptotic behavior of $\boldsymbol{v}_{\widehat{\Theta}} \sim \mathcal{N}(\boldsymbol{v}_{\Theta}, \boldsymbol{CRB}_{\boldsymbol{v}_{\Theta}})$. The theoretical MSE is computed from [Paper, (9)].

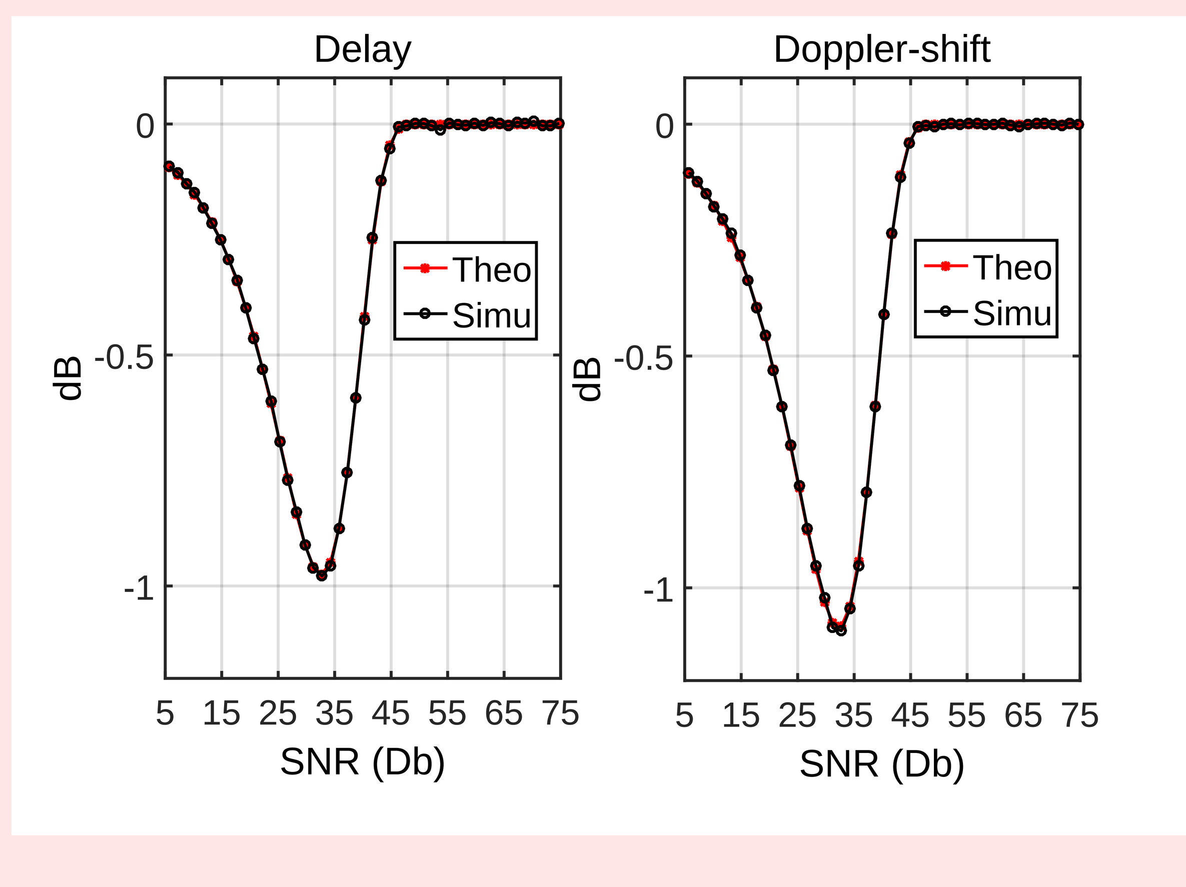


Figure: Average MSE shrinkage factor