Assessment of local influence for the analysis of agreement

Carla Leal[1] **Manuel Galea**[2] **Felipe Osorio**[3]

[1]Universidad Viña del Mar [2]Pontificia Universidad Católica de Chile
[3]Universidad Técnica Federico Santa María

Webinar: Concordance and Covariance functions for Environmental Modelling
July 21, 2020

Biometrical Journal **61**, 955-972 (2020).

Outline

1. Agreement measures
   - Concordance correlation coefficient (CCC)
   - Probability of agreement (PA)
2. A motivating example.
3. Local influence measures.
4. Numerical experiments
   - Monte Carlo simulation study.
   - Application: Transient sleep disorder.
5. Concluding remarks and future work.

Concordance correlation coefficient (Lin, 1989)

Let $(x_{11}, x_{12}), \ldots, (x_{n1}, x_{n2})$ be a bivariate random sample from a population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

A method to quantify the degree of agreement between the variables x_1 and x_2 corresponds to the CCC (Lin, 1989),[1] which is defined as

$$\rho_{\mathsf{c}} = \frac{2\sigma_{12}}{\sigma_{11} + \sigma_{22} + (\mu_1 - \mu_2)^2},$$

where μ_j and σ_{jj} are the mean and variance of the measurements obtained by the jth method or instrument of measurement $(j = 1, 2)$, and σ_{12} is the covariance between the measurements from methods 1 and 2.

It is easy to see that the CCC can be written as

$$\rho_{\mathsf{c}} = \rho_{12} C_{12}, \qquad C_{12} = 2\left[\frac{\sqrt{\sigma_{11}\sigma_{22}}}{\sigma_{11} + \sigma_{22} + (\mu_1 - \mu_2)^2}\right].$$

Moreover, a nice property of CCC is $-1 \leq \rho_{\mathsf{c}} \leq 1$.

[1] Biometrics **45**, 225-268

Probability of agreement (Stevens et al., 2017)

Let $D_i = x_{i1} - x_{i2}$ for $i = 1, \dots, n$, be the differences between the measurements obtained by the two instruments. Stevens et al. (2017)[2] introduced the probability of agreement defined as:

$$\psi_{\mathsf{c}} = \mathrm{P}(|D_i| \leq c), \qquad c > 0,$$

where $CAD = (-c, c)$ represents a clinically acceptable difference. Assuming that the observations (x_{i1}, x_{i2}), $i = 1, \dots, n$, were selected from a bivariate normal population yields

$$\psi_{\mathsf{c}} = \Phi\Big(\frac{c - \mu_D}{\sigma_D}\Big) - \Phi\Big(-\frac{c - \mu_D}{\sigma_D}\Big),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal, and $\mu_D = \mu_1 - \mu_2$, $\sigma_D^2 = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$.

[2] Statistical Methods in Medical Research **26**, 2487-2504.

Motivating example (Svetnik et al., 2007)

Svetnik et al. (2007)[3] conducted a clinical study designed to compare the automated and semi-automated scoring of Polysomnographic (PSG) recordings used to diagnose transient sleep disorders.

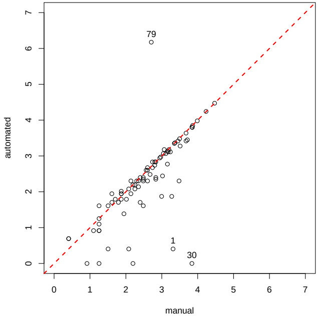
The study considered 82 patients who were given a sleep-inducing drug (Zolpidem 10 mg). Measurements of latency to persistent sleep (LPS: lights out to the beginning of 10 consecutive minutes of uninterrupted sleep) were obtained using six different methods.

We focus on two of these methods: fully manual scoring (Manual) and automated scoring by the Morpheus software (Automatic).

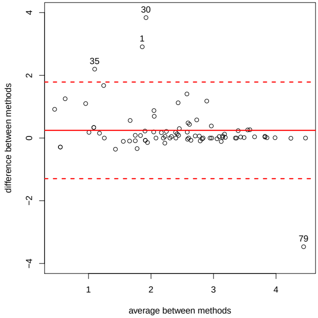
Let $\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top$, for $i = 1, \dots, 82$, be the $\log(\text{LPS})$ measurements obtained with the manual and automatic methods, respectively.

[3] SLEEP **30**, 1562-1574.

Motivating example (Svetnik et al., 2007)



(a)



(b)

return

Sample estimators for ρ_c and ψ_c

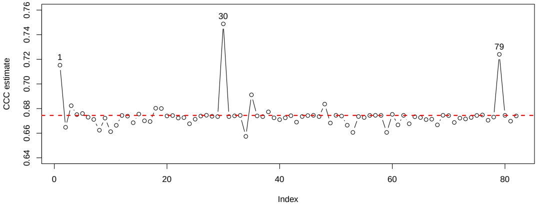
Observations	CCC	$\widehat{\rho}_{12}$	$\widehat{C}_{12}$	CI (95%)
With all subjects[4]	0.674 (0.056)	0.715	0.943	(0.564, 0.785)
Obs. 1,30,79 removed	0.860 (0.028)	0.890	0.967	(0.806, 0.915)

Observations	ψ_c	SE	CI (95%)
With all subjects	0.975	0.034	(0.909, 1.000)
Obs. 1,30,79 removed	0.999	0.001	(0.998, 1.000)

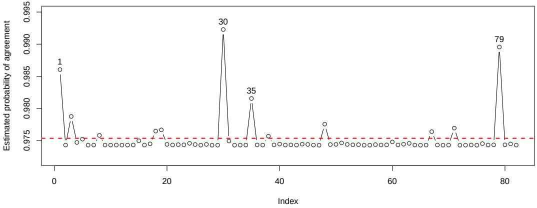
If subjects 1, 30 and 79 are eliminated from the dataset, the degree of agreement is increased by 28% mainly due to an increase in ρ_{12} (also called precision).

[4] McBride (2005) suggested 0.650 as a cutt-off for the CCC.

Motivating example (Svetnik et al., 2007)



(a) case deletion plot for CCC



(b) case deletion plot for ψ_c

To assess the influence of extreme observations on the maximum likelihood estimates, Cook (1986)[5] proposed to study the likelihood displacement

$$LD(\omega) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))\},$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ are the MLE based on the postulated and perturbated models, which are defined as

$$\mathcal{P} = \{p(\boldsymbol{x};\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$$

and,

$$\mathcal{P}_\omega = \{p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\omega}) : \boldsymbol{\theta} \in \Theta, \boldsymbol{\omega} \in \Omega\},$$

respectively, with $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$ satisfying $\mathcal{P}_{\omega_0} = \mathcal{P}$, for a null perturbation, $\boldsymbol{\omega}_0$.

[5] Journal of the Royal Statistical Society, Series B **48**, 133-169

Let $f(\boldsymbol{\omega})$ be a measure of influence. The main aim of the local influence is to analyze the curvature of the influence surface $\varphi(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f(\boldsymbol{\omega}))^\top$ at the critical point $\boldsymbol{\omega}_0$.

Consider $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \varepsilon \boldsymbol{h}$, where $\boldsymbol{h}$ is a unitary direction ($\|\boldsymbol{h}\| = 1$) and $\varepsilon \in \mathbb{R}$. When $f(\boldsymbol{\omega}) = LD(\boldsymbol{\omega})$ its local behavior around $\varepsilon = 0$ for a direction $\boldsymbol{h}$ can be characterized by

$$C_h = \boldsymbol{h}^\top \ddot{\boldsymbol{F}} \boldsymbol{h}, \qquad \ddot{\boldsymbol{F}} = \frac{\partial^2 \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top}\Big|_{\omega=\omega_0}.$$

Moreover, Cook (1986) shows that

$$\ddot{\boldsymbol{F}} = 2\boldsymbol{\Delta}^\top (-\ddot{\boldsymbol{L}})^{-1} \boldsymbol{\Delta},$$

with

$$\ddot{\boldsymbol{L}} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}, \qquad \boldsymbol{\Delta} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top},$$

which must be evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, and $\ell(\boldsymbol{\theta})$ and $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ denote the log-likelihood functions arising from $\mathcal{P}$ and $\mathcal{P}_\omega$.

Local influence diagnostics

For general objective functions, $f(\boldsymbol{\omega})$, we have that (Cook, 1986) the normal curvature assumes the form

$$C_{f,h} = \frac{\boldsymbol{h}^\top \boldsymbol{H}_f \boldsymbol{h}}{(1 + \nabla_f^\top \nabla_f)\boldsymbol{h}^\top (\boldsymbol{I} + \nabla_f \nabla_f^\top)\boldsymbol{h}},$$

where $\nabla_f = \partial f(\boldsymbol{\omega})/\partial \boldsymbol{\omega}\big|_{\omega=\omega_0}$ and $\boldsymbol{H}_f = \partial^2 f(\boldsymbol{\omega})/\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top\big|_{\omega=\omega_0}$, whereas the conformal normal curvature in the direction $\boldsymbol{h}$ evaluated at $\boldsymbol{\omega}_0$ (Poon and Poon, 1999)[6] is given by

$$B_{f,h} = \frac{\boldsymbol{h}^\top \boldsymbol{H}_f \boldsymbol{h}}{\|\boldsymbol{H}_f\|_M \boldsymbol{h}^\top (\boldsymbol{I} + \nabla_f \nabla_f^\top)\boldsymbol{h}}.$$

An interesting property of the conformal curvature is that $0 \leq |B_{f,h}| \leq 1$.

[6] Journal of the Royal Statistical Society, Series B **61**, 51-61

Local influence diagnostics

The first-order approach for local influence (Cadigan and Farrell, 2002)[7] is measured using the directional derivative of $f(\boldsymbol{\omega})$, which is given by

$$S_{f,h} = \frac{\partial f(\boldsymbol{\omega})}{\partial \varepsilon}\Big|_{\varepsilon=0} = \boldsymbol{h}^\top \nabla_f,$$

where $\nabla_f = \partial f(\boldsymbol{\omega})/\partial \boldsymbol{\omega}|_{\omega=\omega_0}$. In the case that $\nabla_f \neq \mathbf{0}$, the direction of the maximum local slope is

$$\boldsymbol{h}_{\max} = \frac{\nabla_f}{\|\nabla_f\|}.$$

Remark:

First-order local influence may be unable to detect some significant directions with large curvature (see Wu and Luo 1993[8] and Cadigan and Farrell, 2002).

[7] Applied Statistics **51**, 469-483.

[8] Journal of the Royal Statistical Society, Series B **55**, 929-936.

To construct influence measures of the first and second order, Zhu et al. (2007)[9] introduced the matrix $\boldsymbol{G}(\boldsymbol{\omega})$ defined as the Fisher information matrix with respect to $\boldsymbol{\omega}$, with elements

$$g_{ij}(\boldsymbol{\omega}) = \mathsf{E}_{\omega}\left\{\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i}\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_j}\right\}, \qquad i,j = 1,\dots,n,$$

where $\mathsf{E}_{\omega}(\cdot)$ indicates that the expectation is taken with respect to the density function $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\omega})$.

The first-order influence measure (FI) in the direction $\boldsymbol{h}$ is given by

$$\mathrm{FI}_{f,h} = \frac{\boldsymbol{h}^{\top}\nabla_f\nabla_f^{\top}\boldsymbol{h}}{\boldsymbol{h}^{\top}\boldsymbol{G}(\boldsymbol{\omega}_0)\boldsymbol{h}},$$

The second-order influence measure (SI) in the direction $\boldsymbol{h}$ is given by

$$\mathrm{SI}_{f,h} = \frac{\boldsymbol{h}^{\top}\tilde{\boldsymbol{H}}_f\boldsymbol{h}}{\boldsymbol{h}^{\top}\boldsymbol{G}(\boldsymbol{\omega}_0)\boldsymbol{h}},$$

where $\boldsymbol{G}(\boldsymbol{\omega}_0)$ is the metric tensor matrix evaluated at $\boldsymbol{\omega}_0$.

[9] The Annals of Statistics **35**, 2565-2588.

In the definition of $\mathrm{SI}_{f,h}$, $\tilde{\boldsymbol{H}}_f$ denotes the covariant Hessian matrix at $\boldsymbol{\omega}_0$, with (i,j)th element given by

$$(\tilde{\boldsymbol{H}}_f)_{ij} = \frac{\partial}{\partial \omega_i}\Big(\frac{\partial f(\boldsymbol{\omega})}{\partial \omega_j}\Big)\Big|_{\omega=\omega_0} - \sum_{s,r} g^{r,s}(\boldsymbol{\omega})\Gamma^0_{ijs}(\boldsymbol{\omega})\Big(\frac{\partial f(\boldsymbol{\omega})}{\partial \omega_r}\Big)\Big|_{\omega=\omega_0},$$

in which $g^{r,s}(\boldsymbol{\omega})$ is the (r,s)th element of $\boldsymbol{G}(\boldsymbol{\omega})^{-1}$ and

$$\Gamma^0_{ijs}(\boldsymbol{\omega}) = \frac{1}{2}\Big\{\frac{\partial}{\partial \omega_i} g(\boldsymbol{\omega})_{js} + \frac{\partial}{\partial \omega_j} g_{is}(\boldsymbol{\omega}) - \frac{\partial}{\partial \omega_s} g_{ij}(\boldsymbol{\omega})\Big\},$$

denotes the Christoffel symbol for the Lévi-Civita connection.

We consider $\widehat{\rho}_c(\boldsymbol{\omega})$ and $\widehat{\psi}_c(\boldsymbol{\omega})$ as objective functions

$$\widehat{\rho}_c(\boldsymbol{\omega}) = \frac{2\widehat{\sigma}_{12}(\boldsymbol{\omega})}{\widehat{\sigma}_{11}(\boldsymbol{\omega}) + \widehat{\sigma}_{22}(\boldsymbol{\omega}) + (\widehat{\mu}_1(\boldsymbol{\omega}) - \widehat{\mu}_2(\boldsymbol{\omega}))^2},$$

and

$$\widehat{\psi}_c(\boldsymbol{\omega}) = \Phi\Big(\frac{c - \widehat{\mu}_D(\boldsymbol{\omega})}{\widehat{\sigma}_D(\boldsymbol{\omega})}\Big) - \Phi\Big(-\frac{c - \widehat{\mu}_D(\boldsymbol{\omega})}{\widehat{\sigma}_D(\boldsymbol{\omega})}\Big),$$

where $\widehat{\mu}_D(\boldsymbol{\omega}) = \widehat{\mu}_1(\boldsymbol{\omega}) - \widehat{\mu}_2(\boldsymbol{\omega})$, $\widehat{\sigma}^2_D(\boldsymbol{\omega}) = \widehat{\sigma}_{11}(\boldsymbol{\omega}) + \widehat{\sigma}_{22}(\boldsymbol{\omega}) - 2\widehat{\sigma}_{12}(\boldsymbol{\omega})$, and

$$\widehat{\mu}_j(\boldsymbol{\omega}) = \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \omega_i x_{ij},$$

$$\widehat{\sigma}_{jk}(\boldsymbol{\omega}) = \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \omega_i (x_{ij} - \widehat{\mu}_j(\boldsymbol{\omega}))(x_{ik} - \widehat{\mu}_k(\boldsymbol{\omega})),$$

for $j, k = 1, 2$, and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$.

The density of the perturbed model, is given by

$$p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\omega}) = \prod_{i=1}^{n}\Big[(2\pi)^{-d/2}|\omega_i^{-1}\boldsymbol{\Sigma}|^{-1/2}\exp\big\{-\tfrac{1}{2}\omega_i(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\big\}\Big].$$

The null perturbation is $\boldsymbol{\omega}_0 = \mathbf{1}_n$ in which case $\mathcal{P}_{\omega_0} = \mathcal{P}$ and $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$.

This yields to the matrix $\boldsymbol{G}(\boldsymbol{\omega}_0) = \boldsymbol{I}_n$, and we verify that the perturbation scheme induced by the model $\mathcal{P}_\omega$ is appropiate.

The first- and second-order influence measures are reduced to

$$\mathrm{FI}_{f,h} = \boldsymbol{h}^\top\nabla_f\nabla_f^\top\boldsymbol{h}, \qquad \text{and} \qquad \mathrm{SI}_{f,h} = \boldsymbol{h}^\top\tilde{\boldsymbol{H}}_f\boldsymbol{h},$$

respectively, for each objective function, either $\widehat{\rho}_c(\boldsymbol{\omega})$ or $\widehat{\psi}_c(\boldsymbol{\omega})$.

The first-order derivative required in $\mathrm{FI}_{\widehat{\rho}_c,h}$, as well as $C_{\widehat{\rho}_c,h}$ and $B_{\widehat{\rho}_c,h}$, assumes the form

$$\nabla_{\widehat{\rho}_c} = \frac{\widehat{\rho}_c}{n\widehat{\sigma}_{12}}\big(\boldsymbol{z}_1 \odot \boldsymbol{z}_2 - \widehat{\sigma}_{12}\mathbf{1}\big) - \frac{\widehat{\rho}_c^2}{2n\widehat{\sigma}_{12}}\boldsymbol{z}_*,$$

where $\boldsymbol{z}_j = (z_{1j}, \ldots, z_{nj})^\top$ with $z_{ij} = z_{ij} - \widehat{\mu}_j$, for $i = 1, \ldots, n$; $j = 1, 2$,

$$\boldsymbol{z}_* = (\boldsymbol{z}_1 \odot \boldsymbol{z}_1 - \widehat{\sigma}_{11}\mathbf{1}_n) + (\boldsymbol{z}_2 \odot \boldsymbol{z}_2 - \widehat{\sigma}_{22}\mathbf{1}_n) + 2(\widehat{\mu}_1 - \widehat{\mu}_2)(\boldsymbol{z}_1 - \boldsymbol{z}_2),$$

and $\odot$ represents the Hadamard product.

Moreover, $\tilde{\boldsymbol{H}}_{\widehat{\rho}_c} = \boldsymbol{H}_{\widehat{\rho}_c} + \mathrm{diag}(\nabla_{\widehat{\rho}_c})$, with

$$\boldsymbol{H}_{\widehat{\rho}_c} = \boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_3.$$

For definition of $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\Gamma}_3$ see Leal et al. (2019).

Furthermore $\nabla_{\widehat{\psi}_c} = \partial \widehat{\psi}_c(\boldsymbol{\omega})/\partial \boldsymbol{\omega}|_{\omega=\omega_0}$ assumes the form

$$\nabla_{\widehat{\psi}_c} = -\frac{2}{\widehat{\sigma}_D^2}\phi\Big(\frac{c-\widehat{\mu}_D}{\widehat{\sigma}_D}\Big)\boldsymbol{s},$$

with

$$\boldsymbol{s} = \widehat{\sigma}_D(\boldsymbol{Z}_1 - \boldsymbol{Z}_2) + \frac{1}{2}\Big(\frac{c-\widehat{\mu}_D}{\widehat{\sigma}_D}\Big)\Big\{\frac{n-2}{n}(\boldsymbol{Z}_1 - \boldsymbol{Z}_2) \odot (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) - \widehat{\sigma}_D^2 \mathbf{1}\Big\},$$

where $\phi(\cdot)$ denotes the density funtion of the standard normal.

$\boldsymbol{H}_{\widehat{\psi}_c} = \partial^2 \widehat{\psi}_c(\boldsymbol{\omega})/\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top|_{\omega=\omega_0}$ can be written as,

$$\boldsymbol{H}_{\widehat{\psi}_c} = 2\phi\Big(\frac{c-\widehat{\mu}_D}{\widehat{\sigma}_D}\Big)\Big\{\boldsymbol{\Delta}_1 - \frac{1}{\widehat{\sigma}_D^2}(\boldsymbol{\Delta}_2 + \boldsymbol{\Delta}_3 + \boldsymbol{\Delta}_4) - \frac{1}{\widehat{\sigma}_D^4}\boldsymbol{s}\boldsymbol{s}^\top\Big\}.$$

Details on the definition of $\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \boldsymbol{\Delta}_3$ and $\boldsymbol{\Delta}_4$ can be found in Leal et al. (2019).

Monte Carlo simulation study

We generate 500 datasets of sample sizes $n = 25, 50, 100$ and 200 from $\mathsf{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}.$$

To introduce an outlier, for each dataset, a single observation of the second variable x_2 was changed to $x_2 + \delta$, where $\delta = 0.5, 1.5, 2.0, 2.5, 3.0$ and 3.5.

We find the unitary direction related to the maximum local slope, normal and conformal curvatures and first- and second-order influence measures for $\widehat{\rho}_c(\boldsymbol{\omega})$ and $\widehat{\psi}(\boldsymbol{\omega})$.

For each δ, the percentages of detecting the outlier were computed using the following threshold:

$$M_j = |\boldsymbol{h}_{\mathsf{max}}|_j > \overline{M} + 2\,\mathsf{sd}(M),$$

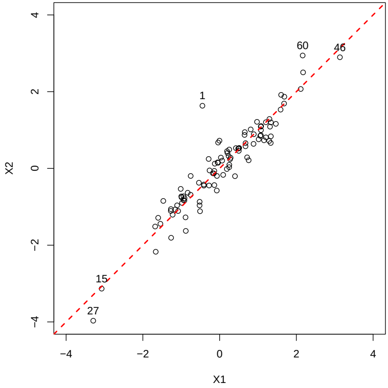
where $\mathsf{sd}(M)$ denotes the standard deviation of M_j, $j = 1, \dots, n$.

All the diagnostic measures described in our work have been implemented in an R code available at github[10]

[10] URL: https://github.com/faosorios/CCC/

Monte Carlo simulation study: A typical dataset

Scatter plot of a typical dataset (with $\delta = 2$) from the simulation experiment:



return

Monte Carlo simulation study

Outlier detection percentage using $\widehat{\rho}_c(\boldsymbol{\omega})$ as objective function

n	Influence measure	δ 0.5	1.0	1.5	2.0	2.5	3.0	3.5
25	C	11.4	33.8	66.4	77.8	86.4	93.4	95.4
	B	11.4	33.8	66.4	77.8	86.4	93.4	95.4
	FI	28.4	74.4	96.0	99.0	99.8	100.0	100.0
	SI	11.4	33.8	66.4	77.8	86.4	93.4	95.4
	FI and SI	6.4	27.4	63.2	76.8	86.2	93.4	95.4
50	C	9.6	34.2	59.4	77.4	90.0	96.4	97.0
	B	9.6	34.2	59.4	77.4	90.0	96.4	97.0
	FI	26.0	76.0	96.4	100.0	100.0	100.0	100.0
	SI	9.6	34.2	59.4	77.4	90.0	96.4	97.0
	FI and SI	5.6	28.4	58.0	77.4	90.0	96.4	97.0
100	C	11.8	30.8	53.8	74.8	92.4	98.4	98.0
	B	11.8	30.8	53.8	74.8	92.4	98.4	98.0
	FI	27.0	77.2	97.2	100.0	100.0	100.0	100.0
	SI	11.8	30.8	53.8	74.8	92.4	98.4	98.0
	FI and SI	8.8	26.8	52.8	74.8	92.4	98.4	98.0
200	C	16.0	42.8	58.2	68.0	89.6	98.2	99.0
	B	16.0	42.8	58.2	68.0	89.6	98.2	99.0
	FI	26.4	79.2	97.2	100.0	100.0	100.0	100.0
	SI	16.0	42.8	58.2	68.0	89.6	98.2	99.0
	FI and SI	9.6	36.6	57.0	68.0	89.6	98.2	99.0

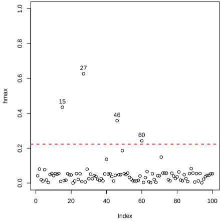
Monte Carlo simulation study

Outlier detection percentage using $\widehat{\psi}_c(\boldsymbol{\omega})^{11}$ as objective function

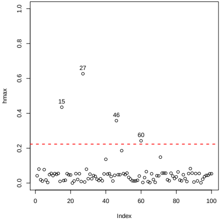
n	Influence measure	δ 0.5	1.0	1.5	2.0	2.5	3.0	3.5
25	C	24.6	77.6	98.2	100.0	100.0	100.0	100.0
	B	24.6	77.6	98.2	100.0	100.0	100.0	100.0
	FI	26.4	81.4	98.6	100.0	100.0	100.0	100.0
	SI	24.6	77.6	98.2	100.0	100.0	100.0	100.0
	FI and SI	24.6	77.4	98.2	100.0	100.0	100.0	100.0
50	C	23.4	74.0	97.8	100.0	100.0	100.0	100.0
	B	23.4	74.0	97.8	100.0	100.0	100.0	100.0
	FI	27.0	81.4	98.2	100.0	100.0	100.0	100.0
	SI	23.4	74.0	97.8	100.0	100.0	100.0	100.0
	FI and SI	23.4	74.0	97.6	100.0	100.0	100.0	100.0
100	C	17.2	58.8	94.6	99.8	100.0	100.0	100.0
	B	17.2	58.8	94.6	99.8	100.0	100.0	100.0
	FI	26.8	82.2	98.4	100.0	100.0	100.0	100.0
	SI	17.2	58.8	94.6	99.8	100.0	100.0	100.0
	FI and SI	16.8	58.2	94.6	99.8	100.0	100.0	100.0
200	C	16.8	59.4	89.2	98.8	100.0	100.0	100.0
	B	16.8	59.4	89.2	98.8	100.0	100.0	100.0
	FI	27.4	83.6	98.8	100.0	100.0	100.0	100.0
	SI	16.8	59.4	89.2	98.8	100.0	100.0	100.0
	FI and SI	15.8	57.4	89.0	98.8	100.0	100.0	100.0

[11] $c = 2$ was used.

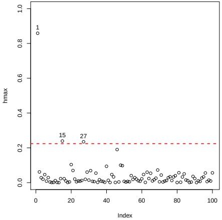
Monte Carlo simulation study: Index plot of $|h_{\max}|$ for $\widehat{\rho}_c(\omega)$ Slide 20



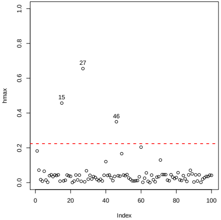
(a) normal curvature, C



(b) conformal curvature, B

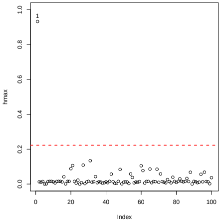


(c) FI

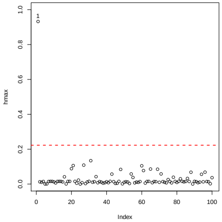


(d) SI

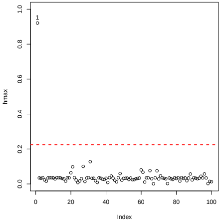
Monte Carlo simulation study: Index plot of $|h_{\max}|$ for $\widehat{\psi}_c(\boldsymbol{\omega})$



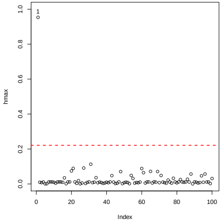
(a) normal curvature, C



(b) conformal curvature, B



(c) FI



(d) SI

Transient sleep disorder (Svetnik et al., 2007)

This dataset was previously analyzed by Feng et al. (2015)[12] using a robust approach within a Bayesian framework.

Figure ▸ Slide 6 reveals that observations 1, 30 and 79 are outside the limits of agreement and therefore can be identified as potential outliers.

Consider the percentage of change of the ML estimates for the fitted model:

Estimate	with all observations	obs. 1,30,79 removed	change (%)
$\widehat{\mu}_1$	2.554	2.526	-1.090
$\widehat{\mu}_2$	2.309	2.313	0.190
$\det(\widehat{\boldsymbol{\Sigma}})$	0.460	0.156	-66.171

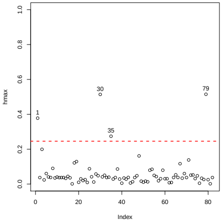
[12] Journal of Biopharmaceutical Statistics **25**, 490-507.

Transient sleep disorder (Svetnik et al., 2007)

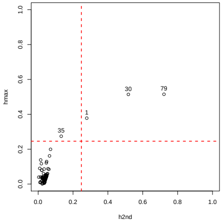
Percentage of change for $\widehat{\rho}_c$, $\widehat{\psi}_c$ and the log-likelihood function:

Observations removed	$\widehat{\rho}_c$	change (%)	$\widehat{\psi}_c$	change (%)	log-likelihood
—	0.674	—	0.975	—	-200.890
1	0.715	6	0.986	1	-192.990
30	0.749	11	0.992	2	-186.448
79	0.724	7	0.990	1	-185.907
1, 30	0.795	18	0.997	2	-175.741
1, 79	0.770	14	0.996	2	-175.982
30, 79	0.808	20	0.999	2	-166.688
1, 30, 79	0.860	28	1.000	3	-150.728

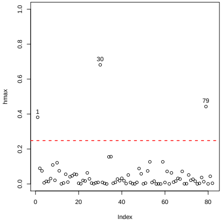
Transient sleep disorder: Index plot of $|\boldsymbol{h}_{\max}|$ for $\widehat{\rho}_c(\boldsymbol{\omega})$



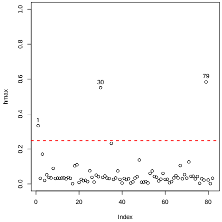
(a) $\boldsymbol{h}_{\max}$, normal curvature



(b) $\boldsymbol{h}_{\max}$ vs. $\boldsymbol{h}_{\text{2nd}}$

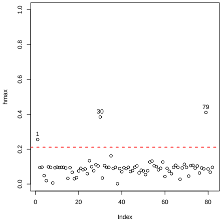


(c) FI

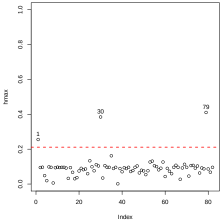


(d) SI

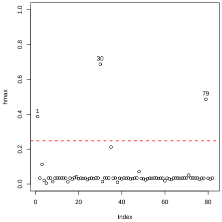
Transient sleep disorder: Index plot of $|h_{\max}|$ for $\widehat{\psi}_c(\omega)$



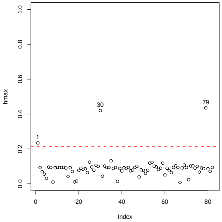
(a) normal curvature, C



(b) conformal curvature, B



(c) FI



(d) SI

Concluding remarks and future work

- Influential data may distort the estimation of the CCC and PA leading to incorrect decisions (replacing one measurement method with another when their degree of agreement is not really true).
- Several diagnostic measures to detect influential data on the estimates of the CCC and PA were proposed.
- A computational implementation of such diagnostic techniques has been made publicly available.
- The empirical results seem to suggest that for our problem, first-order influence measures are efficient for the identification of influential observations.
- Extend the estimation and diagnostics for the CCC and PA considering a multivariate t-distribution.
- Influence diagnostics for the matrix-based concordance correlation coefficient.

EX UMBRA IN SOLEM

Concluding remarks and future work

- Influential data may distort the estimation of the CCC and PA leading to incorrect decisions (replacing one measurement method with another when their degree of agreement is not really true).
- Several diagnostic measures to detect influential data on the estimates of the CCC and PA were proposed.
- A computational implementation of such diagnostic techniques has been made publicly available.
- The empirical results seem to suggest that for our problem, first-order influence measures are efficient for the identification of influential observations.
- Extend the estimation and diagnostics for the CCC and PA considering a multivariate t-distribution.
- Influence diagnostics for the matrix-based concordance correlation coefficient.

References

Cadigan, N., Farrell, P. (2002).
Generalized local influence with applications to fish stock cohort analysis.
Applied Statistics **51**, 469-483.

Cook, R.D. (1986).
Assessment of local influence (with discussion).
Journal of the Royal Statistical Society, Series B **48**, 133-169.

Lin, L. (1989).
A concordance correlation coefficient to evaluate reproducibility.
Biometrics **45**, 225-268.

Leal, C., Galea, M., Osorio, F. (2019).
Assessment of local influence for the analysis of agreement.
Biometrical Journal **61**, 955-972.

Stevens, N.T., Steiner, S.H., MacKay, R.J. (2017).
Assessing agreement between two measurement systems: An alternative to the limits of agreement approach.
Statistical Methods in Medical Research **26**, 2487-2504.

Zhu, H., Ibrahim, J.G., Lee, S., Zhang, H. (2007).
Perturbation selection and influence measures in local influence analysis.
The Annals of Statistics **35**, 2565-2588.