MAT-468: Session 13, Monte Carlo Simulation Studies

Felipe Osorio

http://fosorios.mat.utfsm.cl

Department of Mathematics, UTFSM

Goal:

Explore properties of a statistical procedure using simulation.

In fact, pseudo-random variables can mimic the behavior of (truly) random variables. Consider $T_1, \dots, T_N$ pseudo-random IID variables. Thus,

$$\overline{T} \xrightarrow{\mathsf{P}} \mathsf{E}(T), \qquad s_N^2 \xrightarrow{\mathsf{P}} \mathsf{var}(T), \qquad \frac{\overline{T} - \mathsf{E}(T)}{s_N/\sqrt{N}} \xrightarrow{\mathsf{D}} \mathsf{N}(0,1).$$

We could be interested in essentially any function, for example

$$\mathsf{E}(T^k), \qquad \mathsf{P}(T \leq t) = \mathsf{E}[I(T \leq t)], \qquad M_T(s) = \mathsf{E}[\exp(sT)],$$

and obviously these ideas also apply to random vectors.

Monte Carlo Simulation Studies

In Monte Carlo simulation studies, $T_1, \dots, T_N$ are often statistics, estimators, or test statistics, that are obtained by applying a statistical procedure to simulated data sets, that is

$$T_i = T(X_{i,1}, \dots, T_{i,n}),$$

where $T(\cdot)$ is the function that defines the statistic, and

$$\{X_{i,1}, \dots, X_{i,n}\}, \qquad i = 1, \dots, N,$$

is the ith data set generated **from** a statistical model $\mathcal{P} = \{\mathsf{P}_\theta : \boldsymbol{\theta} \in \Theta\}$.

These methods are often used to evaluate:

- The bias and variance of an estimator.
- Percentiles of a test statistics of pivotal quantity.
- The power function of a hypothesis procedure.
- Coverage probability of a confidence interval.

Next we will see some examples.

We say that a random vector $\boldsymbol{X} = (X_1, \dots, X_p)^\top$ has a multivariate t distribution, with a $\boldsymbol{\mu}$ mean vector, covariance matrix $\boldsymbol{\Sigma}$, and $0 \le \eta < \frac{1}{2}$ shape parameter, if its density function is given by:

$$f(\boldsymbol{x}) = K_p(\eta)|\boldsymbol{\Sigma}|^{-1/2}(1 + c(\eta)D^2)^{-\frac{1}{2\eta}(1+\eta p)}, \qquad \boldsymbol{x} \in \mathbb{R}^p, \tag{1}$$

where

$$K_p(\eta) = \Big(\frac{c(\eta)}{\pi}\Big)^{p/2} \frac{\Gamma(\frac{1+\eta p}{2\eta})}{\Gamma(\frac{1}{2\eta})}, \qquad 0 \le \eta < \tfrac{1}{2},$$

with $c(\eta) = \eta/(1 - 2\eta)$ and $D^2 = (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$.

If a random vector $\boldsymbol{X}$ has a density function (1), we shall denote $\boldsymbol{X} \sim \mathsf{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$[1].

[1] $\eta = 0$ corresponds to $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

Hypothesis in explicit form

We are interested in hypothesis problems of the form:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\lambda}) \qquad \text{vs.} \qquad H_1 : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\lambda}),$$

where $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^k$ is a differentiable function, whose null parametric subspace is given by

$$\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \Lambda \subset \mathbb{R}^m\},$$

and $\partial \boldsymbol{g}(\boldsymbol{\lambda})/\partial \boldsymbol{\lambda}^\top$ being a full-rank matrix.

Wald, score, LR and gradient statistics take the form

$$W = n(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})^\top \boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}), \qquad S = \frac{1}{n}\boldsymbol{U}^\top(\widetilde{\boldsymbol{\theta}})\boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}})\boldsymbol{U}(\widetilde{\boldsymbol{\theta}}),$$

$$LR = 2\{\ell_{\mathsf{o}}(\widehat{\boldsymbol{\theta}}) - \ell_{\mathsf{o}}(\widetilde{\boldsymbol{\theta}})\}, \qquad G = \boldsymbol{U}^\top(\widetilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}).$$

In this context, $\widetilde{\boldsymbol{\theta}} = \boldsymbol{g}(\widetilde{\boldsymbol{\lambda}})$ with $\widetilde{\boldsymbol{\lambda}}$ the restricted MLE of $\boldsymbol{\lambda}$. Under $H_0 : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\lambda})$ the statistics above have a χ^2_{k-m} limiting distribution.

Hypothesis testing regarding the covariance matrix

We are interested in hypothesis written in an explicit form as:

$$H_0 : \boldsymbol{\phi} = \boldsymbol{g}_2(\boldsymbol{\lambda}) \qquad (\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top).$$

- Variance homogeneity of correlated variables: In this case, the intention is to test

  $$H_0 : \sigma_{11} = \sigma_{22} = \cdots = \sigma_{pp},$$

  that can be written in an equivalent form as

  $$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Phi},$$

  where $\sigma^2 > 0$ and $\boldsymbol{\Phi}$ is a $p \times p$ correlation matrix.

- Equicorrelation: the objective is test the hypothesis

  $$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R},$$

  where $\sigma^2 > 0$, $\boldsymbol{R} = (1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p$ with $-1/(p-1) < \rho < 1$.

We are interested in hypothesis written in an explicit form as:

$$H_0 : \boldsymbol{\phi} = \boldsymbol{g}_2(\boldsymbol{\lambda}) \qquad (\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top).$$

- Variance homogeneity of correlated variables: In this case, the intention is to test

  $$H_0 : \sigma_{11} = \sigma_{22} = \cdots = \sigma_{pp},$$

  that can be written in an equivalent form as

  $$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Phi},$$

  where $\sigma^2 > 0$ and $\boldsymbol{\Phi}$ is a $p \times p$ correlation matrix.

- Equicorrelation: the objective is test the hypothesis

  $$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R},$$

  where $\sigma^2 > 0$, $\boldsymbol{R} = (1 - \rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p$ with $-1/(p-1) < \rho < 1$.

Hypothesis testing regarding the covariance matrix

We are interested in hypothesis written in an explicit form as:

$$H_0 : \boldsymbol{\phi} = \boldsymbol{g}_2(\boldsymbol{\lambda}) \qquad (\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \eta)^\top).$$

- Variance homogeneity of correlated variables: In this case, the intention is to test

  $$H_0 : \sigma_{11} = \sigma_{22} = \cdots = \sigma_{pp},$$

  that can be written in an equivalent form as

  $$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Phi},$$

  where $\sigma^2 > 0$ and $\boldsymbol{\Phi}$ is a $p \times p$ correlation matrix.

- Equicorrelation: the objective is test the hypothesis

  $$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R},$$

  where $\sigma^2 > 0$, $\boldsymbol{R} = (1 - \rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p$ with $-1/(p-1) < \rho < 1$.

Monte Carlo simulation study

Based on the simulation study reported by Li, Fang y Zhu (1997, JCGS 6, 435-450).

We generate $M = 10\,000$ datasets of sample sizes $n = 60, 180, 960$ considering:

1) $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and 2) $\mathsf{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$,

for $p = 5, 15, 45$ with

$$\boldsymbol{\mu} = \mathbf{1}, \qquad \boldsymbol{\Sigma} = 0.5\,\boldsymbol{I}_p + 0.5\,\mathbf{1}\mathbf{1}^\top, \quad \eta = 0.25,$$

the equicorrelation hypothesis was tested assuming a normal and multivariate t models

$$H_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}, \qquad \boldsymbol{R} = (1-\rho)\boldsymbol{I} + \rho\mathbf{1}\mathbf{1}^\top,$$

using `MVT` package (routines implemented in C, considering R API 3.3.1).

We ran our simulations on an IBM M4 Server with 2 Intel Xeon E5-2670 processors and 264 GB of RAM (total simulation time: 14 hrs, 29 min, 15 sec).[2]

[2] 360 000 models were fitted.

Results: Empirical sizes, data generated under normal assumption[3]

Test	n	normal fit p = 5	15	45	multivariate t fit p = 5	15	45
LR	60	7.28	22.81	100.00	6.40	21.69	100.00
	180	5.63	8.83	73.07	4.86	7.51	71.16
	960	4.84	5.78	10.83	3.87	3.61	7.31
Wald	60	31.16	99.98	100.00	28.19	99.96	100.00
	180	12.76	79.47	100.00	10.86	75.29	100.00
	960	6.05	15.69	98.78	5.64	12.20	97.20
score	60	5.45	6.57	10.78	4.96	5.76	10.52
	180	5.07	5.73	6.57	4.53	5.08	5.89
	960	4.72	5.31	5.39	4.60	4.69	4.39
gradient	60	5.45	6.57	10.78	5.16	6.03	10.70
	180	5.07	5.73	6.57	4.78	5.33	6.24
	960	4.72	5.31	5.39	4.82	4.87	4.82

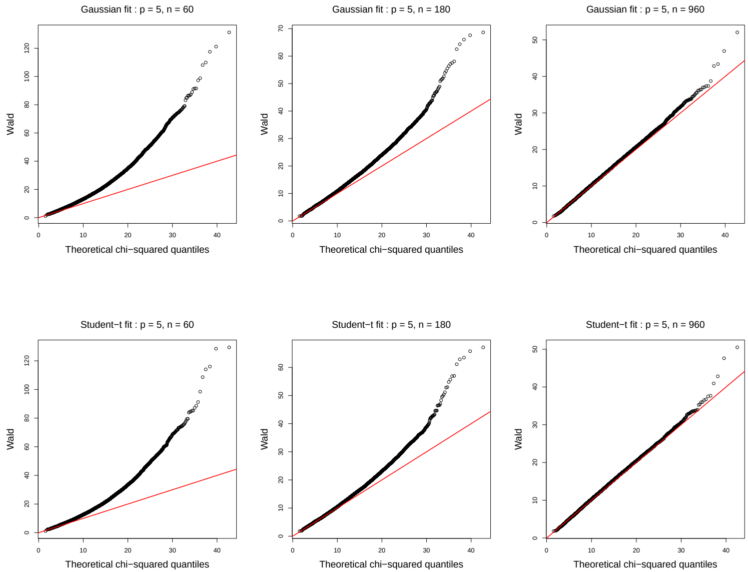
[3] 5% of significance

Results: Empirical sizes, data generated under multivariate t assumption[4]

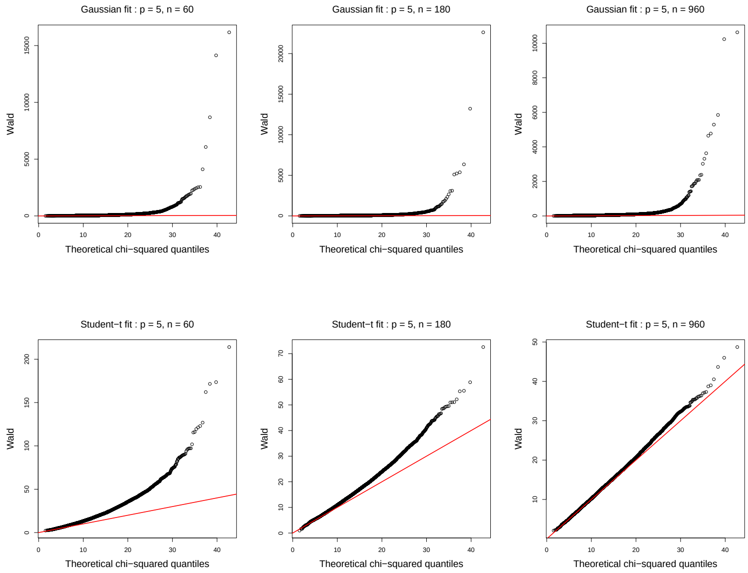
		normal fit			multivariate t fit		
		p			p		
Test	n	5	15	45	5	15	45
LR	60	64.33	99.44	100.00	6.48	23.91	100.00
	180	75.88	99.92	100.00	5.58	8.84	73.05
	960	88.23	100.00	100.00	5.69	5.18	9.50
Wald	60	83.83	100.00	100.00	31.70	99.86	100.00
	180	82.35	100.00	100.00	12.58	72.55	100.00
	960	88.78	100.00	100.00	6.51	13.78	95.14
score	60	61.30	98.44	100.00	4.70	6.31	9.76
	180	75.31	99.90	100.00	4.81	5.49	6.48
	960	88.19	100.00	100.00	5.13	4.88	5.28
gradient	60	61.30	98.44	100.00	4.55	6.24	18.30
	180	75.31	99.90	100.00	4.71	5.01	6.42
	960	88.19	100.00	100.00	5.12	4.70	4.25

[4] 5% of significance

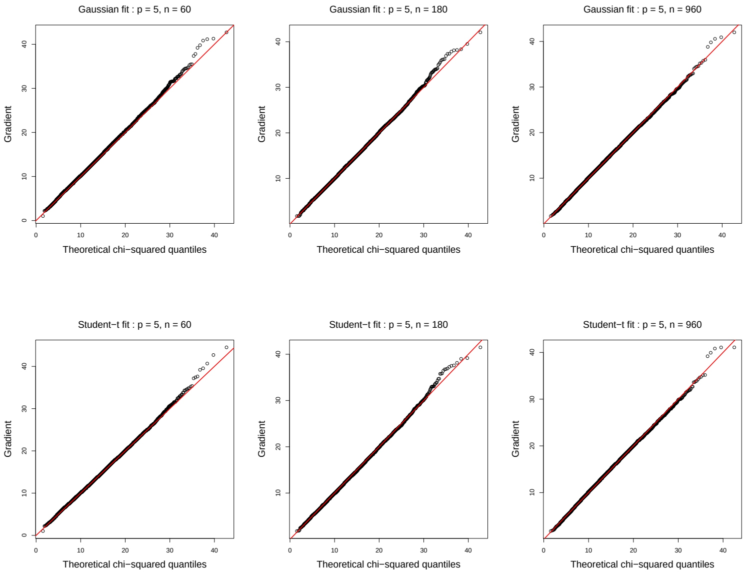
Results: QQ-plot of Wald statistics under normal data, $p = 5$



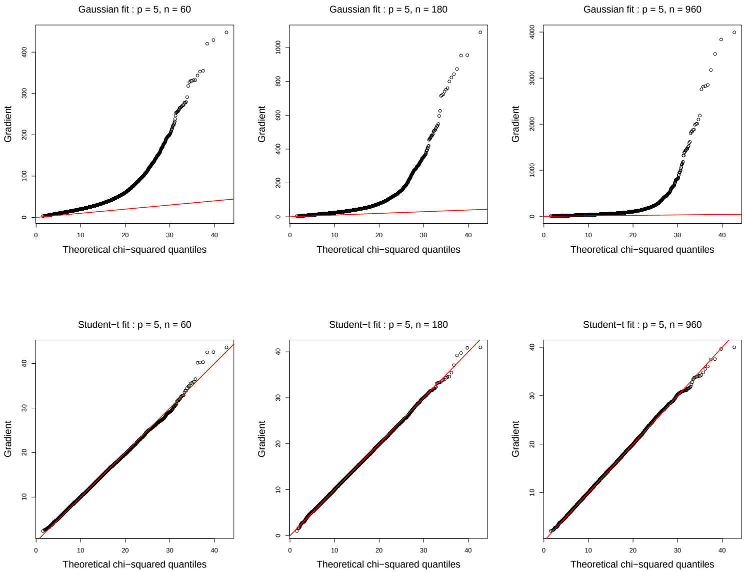
Results: QQ-plot of Wald statistics under multivariate t data, $p = 5$



Results: QQ-plot of gradient statistics under normal data, $p = 5$



Results: QQ-plot of gradient statistics under multivariate t data, $p = 5$



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),[5] 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 $-1 \leq \rho_{\mathsf{c}} \leq 1$.

[5] 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)[6] introduced the probability of agreement defined as:

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

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 $\mu_D = \mu_1 - \mu_2$, $\sigma_D^2 = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$.

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

To assess the influence of extreme observations on the maximum likelihood estimates, Cook (1986)[7] 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, $\boldsymbol{\omega}_0$.

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

Local influence diagnostics

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 $\mathcal{P}_\omega$.

EX UMBRA IN SOLEM

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)[8] 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 $0 \leq |B_{f,h}| \leq 1$.

[8] 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)[9] 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[10] and Cadigan and Farrell, 2002).

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

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

Local influence diagnostics

To construct influence measures of the first and second order, Zhu et al. (2007)[11] 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 $\boldsymbol{\omega}_0$.

[11] 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).

EX UMBRA IN SOLEM

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

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

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).

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:

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

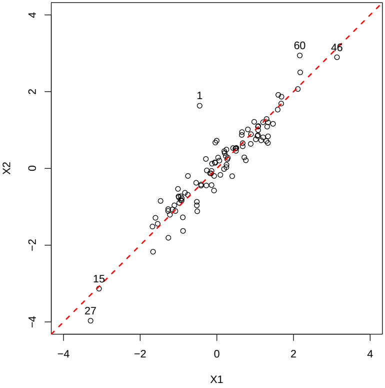
where $\mathsf{sd}(M)$ denotes the standard deviation of M_j, $j = 1, \dots, n$.

All the diagnostic measures described in our work have been implemented in an R code available at github[12]

[12] 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

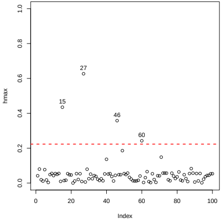
n	Influence measure	δ						
		0.5	1.0	1.5	2.0	2.5	3.0	3.5
25	C	11.4	33.8	66.4	77.8	86.4	93.4	95.4
	B	11.4	33.8	66.4	77.8	86.4	93.4	95.4
	FI	28.4	74.4	96.0	99.0	99.8	100.0	100.0
	SI	11.4	33.8	66.4	77.8	86.4	93.4	95.4
	FI and SI	6.4	27.4	63.2	76.8	86.2	93.4	95.4
50	C	9.6	34.2	59.4	77.4	90.0	96.4	97.0
	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

EX UMBRA IN SOLEM

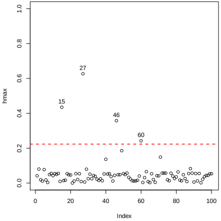
Outlier detection percentage using $\widehat{\psi}_c(\boldsymbol{\omega})$[13] as objective function

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

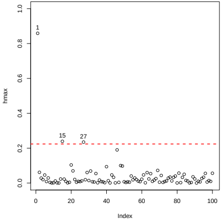
[13] $c = 2$ was used.



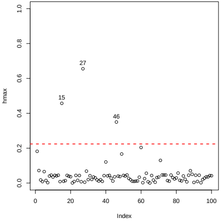
(a) normal curvature, C



(b) conformal curvature, B

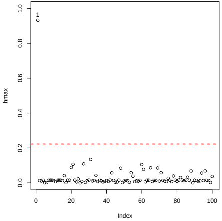


(c) FI

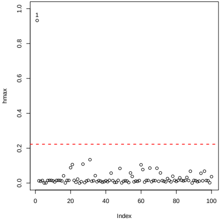


(d) SI

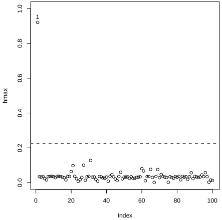
Monte Carlo simulation study: Index plot of $|h_{\max}|$ for $\widehat{\psi}_c(\boldsymbol{\omega})$



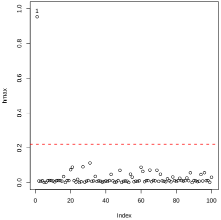
(a) normal curvature, C



(b) conformal curvature, B



(c) FI



(d) SI

Definition 1

An estimator $\widehat{\boldsymbol{\theta}}_n$ is called an extremum estimator if there is an objective function $Q_n(\boldsymbol{\theta})$ such that

$$Q_n(\widehat{\boldsymbol{\theta}}) = \max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}), \qquad \Theta \subseteq \mathbb{R}^p.$$

Remark:

- The notation emphasizes that $Q_n(\boldsymbol{\theta})$ depends on the sample data $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$.
- Proof of the existence and consistency of $\widehat{\boldsymbol{\theta}}_n$ can be found in Gourieroux and Monfort (1995).

Gradient statistic (Terrell, 2002)

Consider $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ random variables with density $f(\cdot;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$, and let

$$\ell_n(\boldsymbol{\theta}) = \log f(\boldsymbol{y};\boldsymbol{\theta}), \qquad \boldsymbol{U}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

be the log-likelihood and score functions, respectively.

It is well known that

$$\frac{1}{\sqrt{n}}\boldsymbol{U}_n(\boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{\mathcal{F}}(\boldsymbol{\theta})),$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta})),$$

where

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = \mathsf{E}\Big\{-\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Big\} = \mathsf{E}\,\{\boldsymbol{U}_n(\boldsymbol{\theta})\boldsymbol{U}_n^\top(\boldsymbol{\theta})\},$$

denotes the Fisher information matrix.

Definition 2 (Terrell, 2002)[14]

The gradient statistic T, to test the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ is given by

$$T = \boldsymbol{U}_n^\top(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

and is asymptotically chi-squared distributed with p degrees of freedom.

Remark:

- If $\ell_n(\boldsymbol{\theta})$ is unimodal, then $T \geq 0$.
- $T \overset{\text{a}}{=} LM$, with
  $$LM = \boldsymbol{U}_n^\top(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{U}_n(\boldsymbol{\theta}_0),$$
  being the Lagrange multiplier test[15] (or score statistic).

[14] Computing Sciences and Statistics **34**, 206-215.

[15] where $\overset{\text{a}}{=}$ denotes asymptotic equivalence.

Problem 1:

Consider the simple hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ with $\boldsymbol{\theta}_0$ fixed. In the context of extreme estimation, we would like to solve

$$\max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}) \qquad \text{subject to:} \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

Problem 2:

We want to test nonlinear hypothesis as $H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, where $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^q$, such that $\boldsymbol{G}(\boldsymbol{\theta}) = \partial \boldsymbol{g}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^\top$ is a $q \times p$ matrix with rank q. In other words, we want to solve the restricted problem:

$$\max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}) \qquad \text{subject to:} \quad \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}.$$

Working assumptions

The bilinear form tests[16] (BF) that we propose are generalizations of Terrell's gradient statistic to the extreme estimation context. We will use the following assumptions:

A1: Let us define $\boldsymbol{A}_n(\boldsymbol{\theta}) = \partial^2 Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ and assume that

$$\boldsymbol{A}_n(\boldsymbol{\theta}) \xrightarrow{\text{a.s.}} \boldsymbol{A},$$

uniformly.

A2: The matrix $\boldsymbol{A}$ is nonsingular.

A3: The secuence

$$\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \xrightarrow{\text{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{B}).$$

Result 1 (Asymptotic normality)

Following Property 24.16 in Gourieroux and Monfort (1995) we know that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\text{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}).$$

[16] Sometimes the term *gradient statistic* is used to indicate LM test for GMM.

Definition 3

The bilinear form statistic to test the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ in extreme estimation, assumes the form:

$$BF = n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{B}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

and under H_0, BF has a $\chi^2(p)$ asymptotic distribution.

Remark:

Assuming that $\boldsymbol{B} = -\boldsymbol{A}$, leads to

$$BF = n\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big)^\top (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

when $Q_n(\boldsymbol{\theta}) = \bar{\ell}_n(\boldsymbol{\theta})$, we obtain the gradient statistic given by Terrell (2002).

Let $\{\widetilde{\boldsymbol{\theta}}_n\}$ be a sequence of estimators that are solution of the problem

$$\max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}) \qquad \text{subject to:} \quad \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}.$$

The resulting constrained estimator $\widetilde{\boldsymbol{\theta}}_n$ satisfies the first-order conditions:

$$\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} - \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}_n)^\top \widetilde{\boldsymbol{\lambda}}_n = \mathbf{0}$$

$$\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}_n) = \mathbf{0},$$

where $\widetilde{\boldsymbol{\lambda}}_n$ is a sequence of Lagrange multipliers.

Problem 2: Bilinear form statistics for testing nonlinear hypothesis

Goal:

We wish to obtain a bilinear form statistic to test nonlinear hypotheses, such as

$$H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0} \qquad \text{against} \qquad H_1 : \boldsymbol{g}(\boldsymbol{\theta}_0) \neq \mathbf{0}.$$

We know that, under H_0,

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{\Omega}), \tag{A.1}$$

where $\boldsymbol{\Omega} = \boldsymbol{G}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{G}^\top$, and

$$\boldsymbol{G} = \frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big|_{\theta=\theta_0}.$$

Problem 2: Bilinear form statistics for testing nonlinear hypothesis

Noticing that

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \stackrel{\mathsf{a}}{=} \boldsymbol{G}(-\boldsymbol{A})^{-1}\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}$$

and from the condition of first order

$$\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \stackrel{\mathsf{a}}{=} \boldsymbol{G}^\top(\widetilde{\boldsymbol{\theta}}_n)\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n.$$

We obtain that

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \stackrel{\mathsf{a}}{=} [\boldsymbol{G}(-\boldsymbol{A})^{-1}\boldsymbol{G}^\top]^{-1}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n). \tag{A.2}$$

Then, using the asymptotic distribution in (A.1), we find

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathsf{D}} \mathsf{N}_q(\boldsymbol{0}, \boldsymbol{S}^{-1}\boldsymbol{\Omega}\boldsymbol{S}^{-1}),$$

where $\boldsymbol{S} = \boldsymbol{G}(-\boldsymbol{A})^{-1}\boldsymbol{G}^\top$.

Let $\boldsymbol{\Omega} = \boldsymbol{R}\boldsymbol{R}^\top$ where $\boldsymbol{R}$ is a nonsingular $q \times q$ matrix. Thus

$$\sqrt{n}\,\boldsymbol{R}^{-1}\boldsymbol{S}\widetilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathsf{D}} \mathsf{N}_q(\boldsymbol{0}, \boldsymbol{I}), \qquad \sqrt{n}\,\boldsymbol{R}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\boldsymbol{0}, \boldsymbol{I}).$$

Result 2

The bilinear form statistic to test the null hypothesis $H_0 : \boldsymbol{g}(\boldsymbol{\theta}_0) = \boldsymbol{0}$ in extremum estimation is given by

$$BF_1 = n\,\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_n),$$

and under H_0, BF_1 has an asymptotic chi-square distribution with q degrees of freedom.

Based on the simulation study introduced by Gregory and Veall (1985)[17].

We generate $M = 5\,000$ datasets of sample sizes $n = 20, 50, 100, 500$ considering the model specification:

$$\boldsymbol{Y} = \boldsymbol{1}_n\theta_1 + \boldsymbol{x}_2\theta_2 + \exp(\boldsymbol{x}_3\theta_3) + \boldsymbol{\epsilon},$$

where $\boldsymbol{x}_j \sim \mathsf{N}_n(\boldsymbol{0}, 0.16\boldsymbol{I})$, $j = 2, 3$ and $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, 0.16\boldsymbol{I})$.

We test two equivalent null hypotheses

$$H_0^A : \theta_2 - \frac{1}{\theta_3} = 0, \qquad H_0^B : \theta_2\theta_3 - 1 = 0.$$

using Bilinear form (BF), Wald (W), Lagrange multiplier (LM) and distance metric (D) statistics.

We ran our simulations on an HP Proliant DL360 Server Intel Xeon E5-2630 processor and 16 GB of RAM (total simulation time: 23 hrs, 40 min, 3 sec).[18]

[17] Econometrica **53**, 1465-1468.

[18] R code available at URL: `https://github.com/faosorios/BF_EE`

Results: Empirical size for a 5% test[19]

Scenario	$(\theta_1, \theta_2, \theta_3)$	n	W^A	W^B	BF^A	BF^B	LM	D
I	(1,10,0.1)	20	0.420	0.176	0.067	0.064	0.084	0.087
		50	0.282	0.106	0.065	0.059	0.068	0.074
		100	0.197	0.077	0.059	0.059	0.061	0.061
		500	0.104	0.052	0.048	0.048	0.049	0.049
II	(1,5,0.2)	20	0.277	0.178	0.068	0.065	0.083	0.086
		50	0.171	0.108	0.058	0.057	0.067	0.068
		100	0.127	0.078	0.058	0.058	0.061	0.062
		500	0.070	0.052	0.047	0.047	0.048	0.048
III	(1,2,0.5)	20	0.145	0.175	0.066	0.067	0.082	0.082
		50	0.096	0.113	0.062	0.057	0.070	0.075
		100	0.078	0.082	0.056	0.056	0.062	0.062
		500	0.049	0.055	0.045	0.045	0.050	0.050
IV	(1,1,1)	20	0.140	0.170	0.084	0.070	0.086	0.101
		50	0.095	0.108	0.062	0.062	0.070	0.070
		100	0.074	0.080	0.066	0.066	0.067	0.067
		500	0.055	0.055	0.056	0.056	0.061	0.061

[19] The superscripts A and B refer to the fact that W and BF are computed using the null hypotheses H_0^A and H_0^B

Monte Carlo simulation study: Power properties of BF

We conducted an additional simulation experiment based on $M = 1\,000$ datasets with sample sizes of $n = 20, 50, 100$, computing the empirical power under the alternative hypotheses:

$$H_1^A : \beta_2 - \delta/\beta_3 = 0,$$

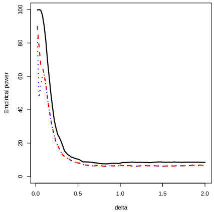
and

$$H_1^B : \beta_2\beta_3 - \delta = 0,$$

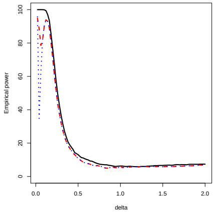
considering $\delta \in [0, 2]$ for each scenario.[20]

[20] Only the empirical power of LM, BF^A and BF^B is reported.

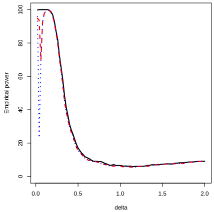
Results: Empirical power of LM, BF^A and BF^B, scenarios I and II



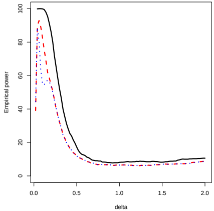
(a) $n = 20$



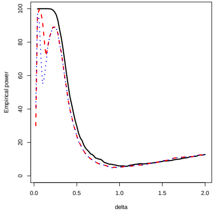
(b) $n = 50$



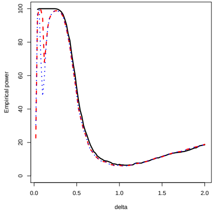
(c) $n = 100$



(d) $n = 20$

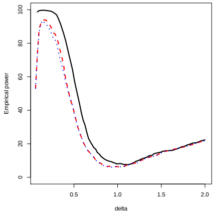


(e) $n = 50$

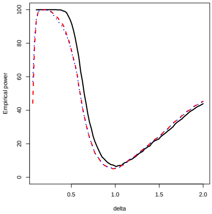


(f) $n = 100$

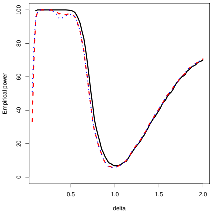
Results: Empirical power of LM, BF^A and BF^B, scenarios III and IV



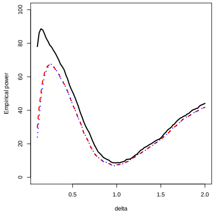
(a) $n = 20$



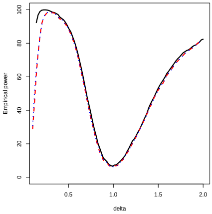
(b) $n = 50$



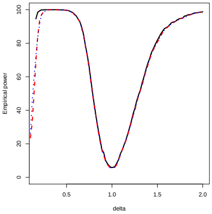
(c) $n = 100$



(d) $n = 20$



(e) $n = 50$



(f) $n = 100$

Estimation of the structural similarity index for image quality assessment

- There are several approaches to study the similarity between two signals, images or processes.
- Image quality assessment aims to quantitatively represent the human perception of quality.
- These indices are designed to study the performance of algorithms for problems such as image compression, image restoration, among others. Full reference algorithms require the distorted **and** reference images.
- The procedure is based on a suitable coefficient that combines the luminance, contrast and structure (correlation) between the images. This type of coefficient has been called Structural Similarity index (SSIM).

Lenna and some distortions of Lenna

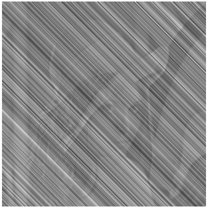
(g) Original image

(h) 8 looks

(i) 32 looks

(j) Speckle noise

(k) Salt-pepper noise



(l) direction $\boldsymbol{h} = (1, 1)$

Mean Squared Error (MSE)

Let $\mathbb{R}_+$ denote the nonnegative real line, and let $\mathbb{R}_+^N$ denote the first orthant. An image is considered an element $\boldsymbol{x} \in \mathbb{R}_+^N$.

If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^N$ are two images, the Mean Squared Error (MSE) between two images $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\mathsf{MSE}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2.$$

Drawbacks:

- The MSE depends strongly on the image scaling.
- The MSE does not represent the Human Vision System (HVS).
- The range for the MSE is the interval $[0, \infty)$.

Distortions of Lenna with same MSE[21]

(a) Original image

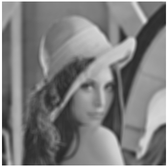
(b) Salt-pepper noise

(c) Gaussian noise

(d) Speckle noise

(e) Mean shift

(f) Contrasting



(g) Blurring

(h) Compression

[21] (a) MSE $= 0$, (b)-(g) MSE $= 225$, and (h) MSE $= 215$.

Structural Similarity Index (SSIM)

Definition (Wang et al., 2004):[22]

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ be two images. The SSIM index is defined as

$$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma},$$

where α, β and γ are non-negative parameters,

$$l(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\,\overline{x}\,\overline{y} + c_1}{\overline{x}^2 + \overline{y}^2 + c_1}, \qquad c(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\,s_x\,s_y + c_2}{s_x^2 + s_y^2 + c_2},$$

$$s(\boldsymbol{x}, \boldsymbol{y}) = \frac{s_{xy} + c_3}{s_x\,s_y + c_3},$$

$\overline{x}$, $\overline{y}$, s_x^2, s_y^2 and s_{xy} represent the sample means, variances and covariance of $\boldsymbol{x}$ and $\boldsymbol{y}$.

The constants c_1, c_2 and c_3 guarantee stability when denominators are close to zero.

[22] IEEE Transactions on Image Processing **13**, 600-612.

Structural Similarity Index (SSIM)

Properties:

- $\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) = 1 \iff \boldsymbol{x} = \boldsymbol{y}$.
- $\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{SSIM}(\boldsymbol{y}, \boldsymbol{x})$.
- SSIM is not a metric, but a pseudo-metric (Brunet et al., 2012).[23]
- $-1 \leq \mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) \leq 1$.
- If $c_1 = c_2 = 0$, we recover the Universal Quality Index (UQI) defined by Wang and Bovik (2002).[24]

[23] IEEE Transactions on Image Processing **21**, 1488-1499.
[24] IEEE Signal Processing Letters **9**, 81-84.

Estimation of α, β and γ

Consider

$$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma},$$

with $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^{\top}$.

Common assumption:

$$\alpha = \beta = \gamma = 1.$$

Our proposal:

- Model-based estimation method for the SSIM parameters.
- A device for testing the hypothesis:

$$H_0 : \alpha = \beta = \gamma = 1.$$

Estimation of the SSIM

- We consider a nonlinear model with multiplicative noise to explain the root mean square error (RMSE). Thus, the following response variable is assumed

$$Z = 1/\,\mathsf{RMSE}(\boldsymbol{x}, \boldsymbol{y}),$$

- The multiplicative nonlinear model adopts the form

$$Z = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma} \cdot e^{u}, \tag{1}$$

  where u is a normal random variable $\mathsf{N}(0, \sigma^2)$.

- We split images $\boldsymbol{x}$ and $\boldsymbol{y}$ into n non-overlapping sub-images denoted as $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$, $i = 1, \ldots, n$. Then, we compute the MSE and SSIM for each pair of sub-images $(\boldsymbol{x}_i, \boldsymbol{y}_i)$.

- The mean and variance functions for model (1) are, respectively, given by

$$\mathsf{E}(Z) = l(\boldsymbol{x},\boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x},\boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x},\boldsymbol{y})^{\gamma} e^{\sigma^2/2}$$
$$\mathsf{var}(Z) = [l(\boldsymbol{x},\boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x},\boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x},\boldsymbol{y})^{\gamma} e^{\sigma^2/2}]^2 (e^{\sigma^2} - 1).$$

- Let $\phi = e^{\sigma^2/2}$ and define

$$f_i(\boldsymbol{\theta}) = \mathrm{SSIM}(\boldsymbol{x}_i, \boldsymbol{y}_i; \boldsymbol{\theta}), \qquad i = 1, \dots, n.$$

  We estimate $\boldsymbol{\psi} = (\boldsymbol{\theta}^\top, \phi)$ using the model

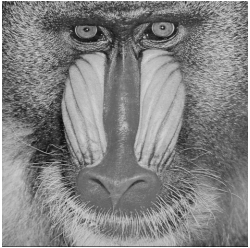
$$Z_i \sim \mathsf{N}(\phi f_i(\boldsymbol{\theta}), f_i^2(\boldsymbol{\theta}) g^2(\phi)), \tag{2}$$

  where $g^2(\phi) = \phi^2(\phi^2 - 1)$.

- Models such as (2) arise frecuently in variance function estimation (Davidian and Carroll, 1987).[25]

[25] Journal of the American Statistical Association **82**, 1079-1091.

(a) texmos2.S512



(b) Baboon

(c) Lenna

Reference images[26] $(\boldsymbol{x} \in \mathbb{R}_+^N)$ were contaminated with multiplicative noise[27] using a Gamma(L, L) distribution, i.e.,

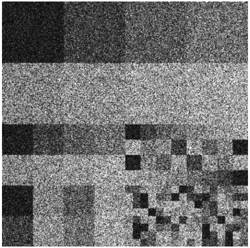
$$\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{w}, \qquad w_t \sim \text{Gamma}(L, L), \qquad t = 1, \ldots, N,$$

for $L = 1, 2, 4, 8, 16$ and 32 looks. For each look 1,000 images were constructed.

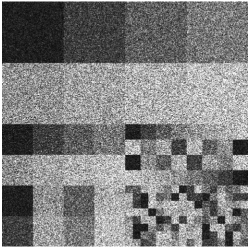
[26] From USC-SIPI image database, URL: http://sipi.usc.edu/database

[27] Available in function `imnoise` from **R** package `SpatialPack`.

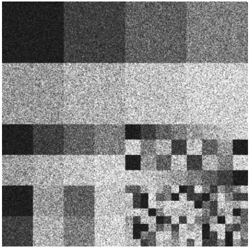
Monte Carlo study: Contamination of texmos2.S512



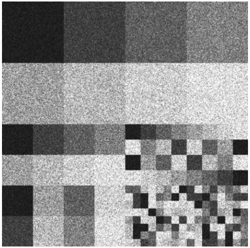
(a) $L = 1$



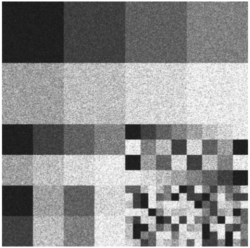
(b) $L = 2$



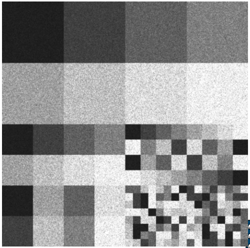
(c) $L = 4$



(d) $L = 8$



(e) $L = 16$



(f) $L = 32$

Number of looks	texmos2.S512		
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
1	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0000
4	1.0389	1.0537	1.0387
8	1.0839	1.1156	1.0836
16	1.1778	1.2381	1.1775
32	1.2668	1.3475	1.2666

Number of looks	Baboon		
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
1	1.0000	1.0000	1.0023
2	1.0010	1.0011	1.0388
4	1.0164	1.0180	1.2089
8	1.1132	1.1231	1.6117
16	1.2674	1.2924	2.2141
32	1.0967	1.1510	2.7799

Number of looks	Lenna		
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
1	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0012
4	1.0032	1.0038	1.0755
8	1.0577	1.0658	1.3928
16	1.2108	1.2371	1.8102
32	1.2960	1.3308	2.1466

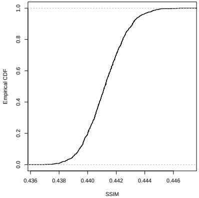
Results: SSIM index under $H_0 : \theta = 1$ and $H_1 : \theta \neq 1$

Number of looks	texmos2.S512 Under H_0	texmos2.S512 Under H_1
1	0.4412	0.4412
2	0.5930	0.5930
4	0.7294	0.7205
8	0.8338	0.8212
16	0.9039	0.8879
32	0.9468	0.9331

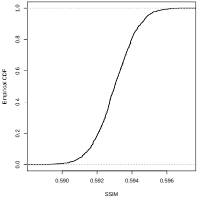
Number of looks	Baboon Under H_0	Baboon Under H_1
1	0.2518	0.2511
2	0.3719	0.3608
4	0.5135	0.4577
8	0.6568	0.5221
16	0.7798	0.5892
32	0.8706	0.6894

Number of looks	Lenna Under H_0	Lenna Under H_1
1	0.3375	0.3375
2	0.4716	0.4713
4	0.6129	0.5934
8	0.7420	0.6659
16	0.8432	0.7387
32	0.9115	0.8219

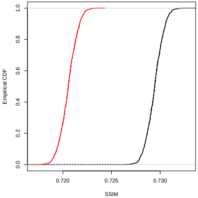
EX UMBRA IN SOLEM



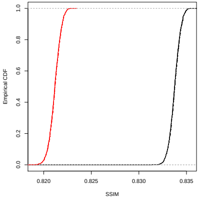
(a) $L = 1$



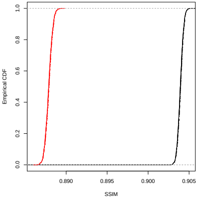
(b) $L = 2$



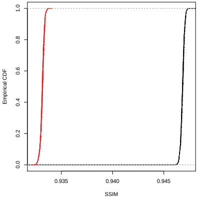
(c) $L = 4$



(d) $L = 8$



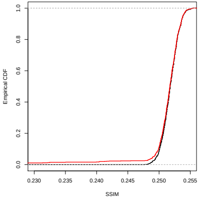
(e) $L = 16$



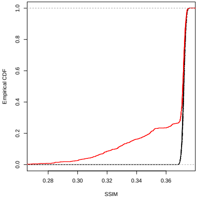
(f) $L = 32$

[28] Empirical CDF of SSIM under H_0 (black) and H_1 (red).

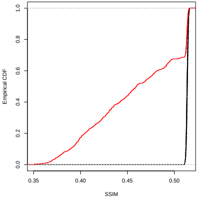
Results: Empirical CDF of SSIM for Baboon[29]



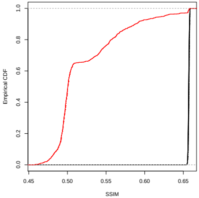
(a) $L = 1$



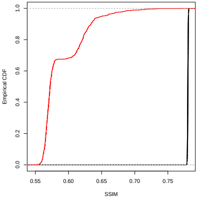
(b) $L = 2$



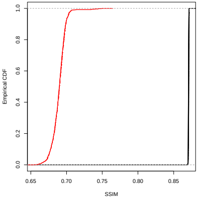
(c) $L = 4$



(d) $L = 8$



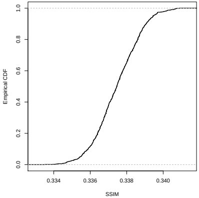
(e) $L = 16$



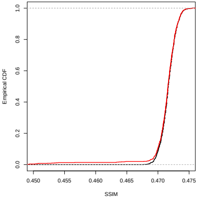
(f) $L = 32$

[29] Empirical CDF of SSIM under H_0 (black) and H_1 (red).

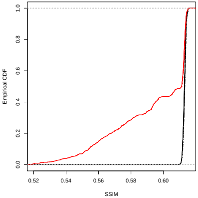
Results: Empirical CDF of SSIM for Lenna[30]



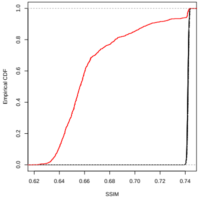
(a) $L = 1$



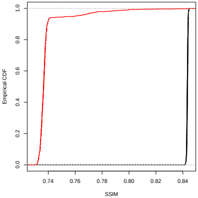
(b) $L = 2$



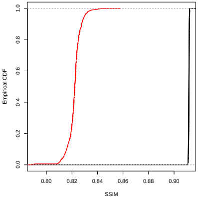
(c) $L = 4$



(d) $L = 8$



(e) $L = 16$



(f) $L = 32$

[30] Empirical CDF of SSIM under H_0 (black) and H_1 (red).