Tests d'adéquation et d'homogénéité basés sur des divergences pondérées

Vlad Stefan BARBU
LMRS, Université de Rouen Normandie, France

Entropies et divergences : modélisation . statistique . algorithmique

LMNO, Université de Caen Normandie, 14-17 mai 2024

Plan

Plan

1. Introduction

Main question

- How can we identify the probabilistic origin of a random phenomenon by focusing, more, on specific characteristics of it (like tail events) ?
- We propose two tests based on weighted divergence measures : a goodness of fit test and a homogeneity test.

Use of divergences in statistics

- perform statistical estimation (for example, MLE)
- perform testing statistical hypothesis
- construct model selection criteria

Presentation based on :

- T. Gkelsinis, A. Karagrigoriou, V. S. Barbu. Statistical inference based on weighted divergence measures with simulations and applications. Statistical Papers, 63, 1511 - 1536, 2022. DOI : https ://doi.org/10.1007/s00362-022-01286-z
- T. Gkelsinis, V. S. Barbu A class of hypothesis tests for general order Markov chains with prior information on the transitions, 2024. under revision.

Plan

2. Divergence measures

A measure of divergence is used as a way to evaluate the distance (divergence) between any two populations or functions. Let p and q be two (probability) density functions which may depend or not on an unknown parameter of fixed finite dimension.

The measures of divergence are not formal distance functions. Any bivariate $D(\cdot,\cdot) \geq 0$ with equality iff its arguments are equal can possibly be used as a measure of information or divergence.

Equally important to the above divergence measures are their limiting versions, known as *divergence rates*. Formally, the divergence rate of a general divergence measure, say D, between two distributions p and q is defined by

$$\lim_{n\to\infty} \frac{1}{n} D(p,q).$$

Some measures of divergence

- Kullback-Leibler (1951)

$$D_{KL}(p,q) = \int p(x)\log(p(x)/q(x))d\mu(x)$$

- Cressie and Read (1984)

$$D_{CR}(p,q) = \frac{1}{\alpha(\alpha-1)}\int \left(p^{\alpha}(x)q^{1-\alpha}(x) - q(x)\right) d\mu(x), \alpha \in \mathbb{R},$$

  where, for $\alpha = 0$ and 1, it is defined by continuity. Note that KL-divergence is obtained for $\alpha \to 0$ and $\alpha \to 1$.

- Liese and Vajda's generalization of of Rényi's divergence

$$D_{R,\alpha}(p,q) = \frac{1}{\alpha(\alpha-1)}\log\left(\int p(x)^{\alpha}q(x)^{1-\alpha}d\mu(x)\right), \ \alpha \neq 0, 1.$$

- Csiszar's φ-divergence (Csiszár, 1963 ; Ali and Silvey, 1966)

$$D_\varphi(p,q) = \int q(x)\varphi\left(\frac{p(x)}{q(x)}\right) d\mu(x),$$

where $\varphi(x)$ is a continuous, differentiable and convex function for $x \geq 0$.

- $\varphi(u) = u \log u$ **Kullback-Leibler** measure
- $\varphi(u) = \frac{1}{2}(1-u)^2$ **Pearson's chi-square** or **Kagan**'s divergence (Kagan, 1963)
- $\varphi(u) = (1-\sqrt{u})^2$ **Matusita**'s divergence (Matusita, 1967) ; $D_{Hellinger}(p,q) = \sqrt{D_{Matusita}(p,q)}$
- $\varphi(u) = \left(u^{\alpha-1} - 1\right)/(\alpha(\alpha-1))$, **Cressie and Read** power divergence (1984)

Weighted divergences

The notion of weighted divergences :

- Introduced in Beliş and Guiaşu (1968), developed in Guiaşu (1971), Sharma et. al (1978), Taneja and Tuteja (1984, 1986), Kapur (1994), Di Crescenzo and Longobardi (2006), Suhov et. al (2016), etc.

For $\mathbf{p} = (p_1, \dots, p_n)$ and $\mathbf{q} = (q_1, \dots, q_n)$ two finite probability distributions and $\mathbf{w} = (w_1, \dots, w_n)$ a vector of weights, $w_i > 0, i = 1, \dots, n$, we define :

- the weighted Shannon entropy measure of $\mathbf{p}$

$$I^S(\mathbf{p}; \mathbf{w}) = -\sum_{i=1}^{n} w_i p_i \log(p_i); \tag{1}$$

- the weighted Shannon divergence (Taneja and Tuteja, 1984)

$$D_S(\mathbf{p}, \mathbf{q}; \mathbf{w}) = \sum_{i=1}^{n} w_i p_i \log\left(\frac{p_i}{q_i}\right). \tag{2}$$

Corrected weighted divergence (Kapur, 1994)

A measure $D(\mathbf{p}, \mathbf{q}; \mathbf{w})$ is said to be an appropriate measure of weighted divergence if the following axioms are fulfilled :

1. It is a continuous function of $(p_1, \ldots, p_n)$, $(q_1, \ldots, q_n)$ and $(w_1, \ldots, w_n)$.
2. It is permutationally symmetric, i.e. it does not change when the triplets (p_1, q_1, w_1), $(p_2, q_2, w_2), \ldots,$ (p_n, q_n, w_n) are permuted among themselves.
3. It is always non-negative and vanishes when $p_i = q_i$ for all $i = 1, \ldots, n$.
4. It is a convex function of $(p_1, \ldots, p_n)$, which has its minimum value zero when $p_i = q_i$ for all $i = 1, \ldots, n$.
5. It reduces to a positive multiple of an ordinary measure of weighted divergence when all the weights are equal.

The corrected weighted Shannon divergence measure corresponding to the Kullback-Leibler measure is given by (Kapur, 1994)

$$D^{W}_{CKL}(\mathbf{p},\mathbf{q};\mathbf{w}) = \sum_{i=1}^{n} w_i \left[p_i \log\left(\frac{p_i}{q_i}\right) - p_i + q_i \right]. \quad (3)$$

Note that the corrected weighted Shannon divergence measure given above verifies Condition 3 of the Kapur's definition ; indeed, one has to factorize by p_i the factor $\left[p_i \log\left(\frac{p_i}{q_i}\right) - p_i + q_i\right]$ in (3) and then to study the sign of the function $f(t) := t\log(t) - t + 1$.

CWKL divergence

Definition (Gkelsinis, T. & Karagrigoriou, A. 2020)

CWKL Divergence

Consider two absolutely continuous probability measures $F_X << \mu$ and $F_Y << \mu$. The CWKL divergence measure between $F_X,\ F_Y$ is defined by

$$D^W_{CKL}(F_X, F_Y) = \frac{\sum_{i=1}^K w_i \left(\int_{S_\mathcal{X}} \left(f_X(x) \log\left(\frac{f_X(x)}{f_Y(x)}\right) - f_X(x) + f_Y(x) \right) d\mu_{|A_i}(x) \right)}{\sum_{l=1}^K w_l \left(\int_{S_X} f_X(x) d\mu_{|A_l}(x) \right)}$$

where $A_i \in \mathcal{A}$ is the i^{th} element of the partition of the support $S_\mathcal{X}$, i.e., $\bigcup_{i=1}^K A_i = S_\mathcal{X},\ A_i \cap A_j = \emptyset\ \forall\ i \neq j,\ \mu_{|A_i}$ is the restricted measure on the subset A_i and $w_i,\ i = 1, \ldots, K$ are the weights directly proportional to the importance of each subset A_i of the support.

Plan

Plan

3. Weighted goodness of fit test (WGoF)

- $X_1, \ldots, X_n \overset{i.i.d}{\sim} F_X$
- $H_0 : F_X \in \mathcal{F}_0$ vs $H_1 : F_X \notin \mathcal{F}_0$
- Several tests according to the statistic :
  - Based on E.D.F. : Kolmogorov-Smirnov $T_n = \sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)|$
  - Based on Order Statistics : Shapiro-Wilk
  - Based on Weighted divergence measures : Weighted Goodness of Fit test (WGoF)

Framework

- Consider a random variable $X \sim F_X$.
- Test the hypothesis : $H_0 : F_X = F^0$, where F^0 is a hypothesized distribution.
- Partition the range of distribution in K classes, say $C_1, \ldots, C_K$.
- The probability of falling into the class i is $P(X \in C_i) = p_i$ where $i = 1, \ldots, K$ and w_i, $i = 1, \ldots, K$ is the weight or the importance of each class.
- Suppose a random sample $X_1, \ldots, X_n$ from the distribution F_X and $\boldsymbol{N} = (N_1, \ldots, N_K)^T$ is the observed number of values falling on each class.

Framework

- It is straight forward that the vector $\boldsymbol{N}$ has a multinomial distribution with parameters $(n, p_1, \ldots, p_K),\ n = \sum_i N_i$.
- Estimator of the probabilities $\boldsymbol{P} = (p_1, \ldots, p_K)^T$ is $\hat{\boldsymbol{P}} = (\hat{p_1}, \ldots, \hat{p_K})^T$, where $\hat{p}_i = \frac{N_i}{n}$, $i = 1, \ldots, K$.
- New hypothesis $H_0 : P = P^0$, where $\boldsymbol{P^0}$ is the hypothesized distribution constructed from the calculation of the probabilities $p_{i0} = P(X \in C_i | F_X = F_0),\ i = 1, ..., K$, under H_0.
- For sufficiently large values of $D^w_{CKL}(\hat{P}, P^0)$ we have to reject the null hypothesis.

Asymptotics of the T_{CWKL} statistics

The theorem below provides the asymptotic distribution of T_{CWKL} which is a natural extension of the result of Frank et al (1998).

Theorem (1)

Assume the CWKL divergence $D^w_{CKL}(\boldsymbol{P}, \boldsymbol{P^0})$ *and its estimator* $D^w_{CKL}(\hat{\boldsymbol{P}}, \boldsymbol{P^0})$*. Under the null hypothesis* $H_0 : \boldsymbol{P} = \boldsymbol{P^0}$ *we have :*

$$T_{CWKL} = 2nD^w_{CKL}(\hat{\boldsymbol{P}}, \boldsymbol{P^0}) \xrightarrow[n\to\infty]{\mathcal{D}} \sum_{i=1}^{r} \beta_i Z_i^2,$$

where Z_i, $i = 1, \ldots, r$ *are iid Standard Normal variables,* β_i, $i = 1, \ldots, r$ *are the eigenvalues of the matrix* $C\Sigma_{\boldsymbol{P^0}}$ *where* $C = (c_{ij})_{i,j=1,\ldots,K} = \begin{cases} 0, & \text{if } i \neq j \\ \frac{w_i}{p_{i0} \sum_{l=1}^{K} w_l p_{l0}}, & \text{if } i = j \end{cases}$,

$\Sigma_{\boldsymbol{P^0}} = \text{diag}(\boldsymbol{P^0}) - \boldsymbol{P^0}(\boldsymbol{P^0})^T$ *and* $r = \text{rank}(\Sigma_{\boldsymbol{P^0}} C \Sigma_{\boldsymbol{P^0}})$.

Test function

- $H_0 : \boldsymbol{P} = \boldsymbol{P^0} \;\; vs \;\; H_1 : \boldsymbol{P} \neq \boldsymbol{P^0}$

- $\phi(\hat{p}_1, ..., \hat{p}_K) = \begin{cases} 1, & T_{CWKL} > q_{1-\alpha} \\ 0, & \text{otherwise} \end{cases}$

  where $q_{1-\alpha}$ is the critical value verifying

$$P(\sum_{i=1}^{r} \beta_i Z_i^2 \leq q_{1-\alpha}) = 1 - \alpha.$$

Consistency

Theorem (2)

The T_{CWKL} test statistics is consistent, in terms of Fraser consistency. That is for every alternative hypothesis $\boldsymbol{P} = \boldsymbol{P^1} \neq \boldsymbol{P^0}$,

$$\lim_{n\to\infty} \beta(\boldsymbol{P^1}) = 1,$$

where $\beta(\boldsymbol{P^1}) = P(T_{CWKL} > q_{1-\alpha}|\boldsymbol{P} = \boldsymbol{P^1})$, is the power function.

Asymptotic distribution under the alternative hypothesis

Theorem (3)

Under the alternative hypothesis $\boldsymbol{P} = \boldsymbol{P}^{\mathbf{1}} \neq \boldsymbol{P}^0$

$$\sqrt{n}\left(D^W_{CKL}(\hat{\boldsymbol{P}}_X, \boldsymbol{P}^0) - D^W_{CKL}(\boldsymbol{P}^{\mathbf{1}}, \boldsymbol{P}^0)\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2(\boldsymbol{w}, \boldsymbol{P}^{\mathbf{1}})),$$

where $\sigma^2(\boldsymbol{w}, \boldsymbol{P}^{\mathbf{1}}) = \nabla D^W_{CKL}(\boldsymbol{P}^{\mathbf{1}}, \boldsymbol{P}^0)^T \Sigma_{\boldsymbol{P}^{\mathbf{1}}} \nabla D^W_{CKL}(\boldsymbol{P}^{\mathbf{1}}, \boldsymbol{P}^0)$ *and* $\Sigma_{\boldsymbol{P}^{\mathbf{1}}} = diag(\boldsymbol{P}^{\mathbf{1}}) - \boldsymbol{P}^{\mathbf{1}}(\boldsymbol{P}^{\mathbf{1}})^T$.

Power Function

Remarque

Using Theorem (3) we derive the power function in an explicit form. In particular, the power function under the alternative hypothesis $\boldsymbol{P^1} \neq \boldsymbol{P}^0$ *is given by,*

$$\beta(\boldsymbol{w}, \boldsymbol{P^1}) = 1 - \Phi_n \left(\frac{q_{1-\alpha} - 2nD^{W}_{CKL}(\boldsymbol{P^1}, \boldsymbol{P}^0)}{2\sqrt{n}\sigma(\boldsymbol{w}, \boldsymbol{P^1})} \right),$$

where unif $\lim_{n\to\infty} \Phi_n = \Phi$ *and* $q_{1-\alpha}$ *is the critical point obtained from the probability of type-1 error.*

Plan

Simulations

- Critical points estimated through Monte-Carlo Simulations (10^4 iterations).
- Power estimation :
  1. For the symmetric case the standard Normal is the hypothesized distribution and the samples raised from various Student's t and Cauchy distributions.
  2. For the non-symmetric case the Log-Normal is the hypothesized distribution and the samples raised from various Weibull, Gamma and Inverse Gaussian distributions.
- $\alpha = 0.05$.

Size of WGoF

Table – Size of WGoF in standard Normal case

standard Normal case			
Partition : $(-\infty,\ -2) \cup (-2,\ 2) \cup (2,\ \infty)$			
n	WGoF with $\boldsymbol{w} = (1, 1, 1)^T$	WGoF with $\boldsymbol{w} = (2, 1, 2)^T$	WGoF with $\boldsymbol{w} = (1, 2, 1)^T$
25	0.0257	0.0285	0.0411
50	0.0289	0.0276	0.0281
100	0.0438	0.0446	0.0420
200	0.0580	0.0581	0.0461
500	0.0528	0.0512	0.0503
1000	0.0536	0.0520	0.0466

Table – Size of WGoF in Log-Normal case

Log-Normal case $(\mu = 0,\ \sigma = 1)$			
Partition : $(0,\ 1) \cup (1,\ 2) \cup (2,\ \infty)$			
n	WGoF with $\boldsymbol{w} = (1, 1, 1)^T$	WGoF with $\boldsymbol{w} = (2, 1, 2)^T$	WGoF with $\boldsymbol{w} = (1, 2, 1)^T$
25	0.0483	0.0541	0.0521
50	0.0485	0.0534	0.0493
100	0.0451	0.0549	0.0511
200	0.0463	0.0488	0.0503
500	0.0461	0.0471	0.0463
1000	0.0499	0.0508	0.0478

Power of WGoF (1/3)

Table – Power comparison of WGoF test in symmetric case

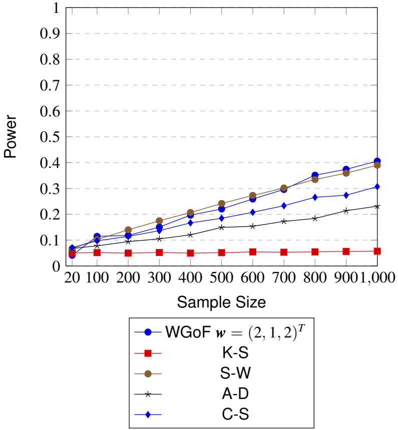
$H_0 : F = F^0 = \mathcal{N}(0,1)$ vs $H_1 : F = t_p$							
Partition : $(-\infty,\ -2) \cup (-2,\ 2) \cup (2,\ \infty)$							
p	n	WGoF with $\boldsymbol{w} = (1,1,1)^T$	WGoF with $\boldsymbol{w} = (2,1,2)^T$	Shapiro-Wilk	Kolmogorov-Smirnov	Anderson-Darling	Chen-Shapiro
5	25	0.1779	0.1232	0.2253	0.0566	0.1958	0.2150
	50	0.3064	0.3147	0.3551	0.0636	0.3057	0.3444
	100	0.5104	0.5229	0.5586	0.0674	0.4824	0.5270
	200	0.8072	0.7936	0.8139	0.0841	0.7339	0.7790
	500	0.9948	0.9929	0.9908	0.1825	0.9813	0.9852
	1000	1	1	1	0.4820	0.9999	1
10	25	0.0817	0.0862	0.1059	0.0560	0.0910	0.1090
	50	0.1267	0.1248	0.1573	0.0533	0.1161	0.1506
	100	0.1914	0.1937	0.2283	0.0490	0.1610	0.2014
	200	0.3166	0.3169	0.3587	0.0579	0.2417	0.3088
	500	0.6668	0.6794	0.6533	0.0645	0.4789	0.6036
	1000	0.9258	0.9314	0.9033	0.0902	0.7767	0.8660
20	25	0.0542	0.0193	0.0767	0.0531	0.0658	0.0756
	50	0.0680	0.0629	0.0826	0.0487	0.0660	0.0872
	100	0.0857	0.0860	0.1065	0.0468	0.0775	0.0958
	200	0.1110	0.1070	0.1476	0.0476	0.0936	0.1186
	500	0.2174	0.2028	0.2391	0.0519	0.1418	0.1866
	1000	0.4084	0.4051	0.3847	0.0559	0.2305	0.3164
30	25	0.0393	0.0440	0.0670	0.0489	0.0593	0.0598
	50	0.0573	0.0523	0.0747	0.0504	0.0611	0.0660
	100	0.0650	0.0646	0.0784	0.0472	0.0679	0.0712
	200	0.0819	0.0841	0.1005	0.0441	0.0695	0.0834
	500	0.1314	0.1305	0.1396	0.0485	0.0902	0.1144
	1000	0.2106	0.2080	0.2104	0.0512	0.1152	0.1630
$H_0 : F = F^0 = \mathcal{N}(0,1)$ vs $H_1 : F = Cauchy(0,p)$							
1	25	0.9179	0.9088	0.9042	0.2405	0.9341	0.9124
	50	0.9802	0.9972	0.9965	0.4734	0.9970	0.9948
	100	1	1	1	0.8568	1	1
	200	1	1	1	0.9996	1	1
	500	1	1	1	1	1	1
	1000	1	1	1	1	1	1
0.5	25	0.9454	0.9584	0.9291	0.0768	0.9346	0.9138
	50	0.9985	1	0.9972	0.1271	0.9967	0.9960
	100	1	1	1	0.2412	1	1
	200	1	1	1	0.5405	1	1
	500	1	1	1	0.9954	1	1
	1000	1	1	1	1	1	1

Power of WGoF (2/3)

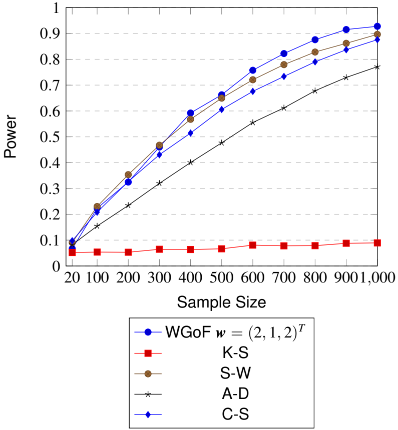
Table – Power comparison of WGoF test in non symmetric case

$H_0 : F = F^0 = Log - Normal(0,1)$ vs $H_1 : F$					
Partition : $(0,\ 2) \cup (2,\ 5) \cup (5,\ \infty)$					
F	n	WGoF with $\boldsymbol{w} = (1,1,1)^T$	WGoF with $\boldsymbol{w} = (1,2,5)^T$	Kolmogorov-Smirnov	Cramer-von Mises
Weibull(1, 1.4)	25	0.0549	0.0672	0.0912	0.0953
	50	0.1386	0.2449	0.1446	0.1323
	100	0.2053	0.2639	0.2334	0.2282
	200	0.3694	0.5273	0.5327	0.4494
	500	0.7644	0.8617	0.9771	0.9307
	1000	0.9804	0.9908	1	0.9998
Gamma(1, 1.4)	25	0.0622	0.0501	0.0918	0.1013
	50	0.1712	0.2436	0.1349	0.1388
	100	0.2320	0.2552	0.2455	0.2253
	200	0.4280	0.5247	0.5294	0.4537
	500	0.7877	0.8570	0.9776	0.9350
	1000	0.9824	0.9894	1	0.9998
IG(3, 1)	25	0.3587	0.4014	0.2175	0.2308
	50	0.6092	0.6318	0.3916	0.4088
	100	0.9069	0.9138	0.6678	0.6926
	200	0.9964	0.9970	0.9520	0.9488
	500	1	1	1	1
	1000	1	1	1	1

Power of WGoF (3/3)



(a) WGoF test power comparison between $\mathcal{N}(0, 1)$ and t_{20}



(b) WGoF test power comparison between $\mathcal{N}(0, 1)$ and t_{10}

Figure – Power comparison of WGoF test

Choice of weights

Several possible choices :

1 Focus on the greater dissimilarities,

$$w_i = \frac{|p_i^0 - \hat{p}_i|}{\kappa} \;\; \forall\, i \in 1, \ldots, K,$$

where $\kappa = \sum_{i=1}^{K} |p_i^0 - \hat{p}_i|$.

2 Perform a test of fit for one tailed distributions with greater significance as we move deeper in the tail ; useful in finance, insurance, extreme value. To emphasize on the tail, a meaningful procedure is to chose the weights

$$w_i = S^{-1}(o_i) \;\; \forall\, i \in 1, \ldots, K,$$

where $S(\cdot)$ is the survival probability and o_i the middle of each of the first $K - 1$ subintervals, where for the last one we can put o_K to be an arbitrary point.

Choice of weights

3 By maximizing the power function $\beta(\boldsymbol{w}, \boldsymbol{P^1})$, with respect to the set of the weights, under a fixed alternative hypothesis. That is, for a fixed alternative $\boldsymbol{P} = \boldsymbol{P^1}$:

$$\boldsymbol{w}_{max} = arg \sup_{\boldsymbol{w} \in \mathbb{R}^K} \gamma(\boldsymbol{w}, \boldsymbol{P^1})$$

4 For someone who wishes to focus on specific subsets of the state space, then, according to the target, the weights can be chosen according to the prior/expert beliefs.

5 When there is no specific interest, the weights can be equal.

Plan

Plan

4. Weighted test of homogeneity (WHom)

- $X_1, \ldots, X_n \overset{i.i.d}{\sim} F_X$ and $Y_1, \ldots, Y_m \overset{i.i.d}{\sim} F_Y$
- $H_0 : F_X = F_Y$ vs $H_1 : F_X \neq F_Y$
- Tests based on divergence measures : Chi-squared test of homogeneity, Weighted test of Homogeneity (WHom)

Framework

- Suppose that we have two independent random samples from unknown distributions $X_1, ..., X_n \sim F_X$, $Y_1, ..., Y_m \sim F_Y$ and we want to perform the following test :

$$H_0 : F_X = F_Y \quad vs \quad H_1 : F_X \neq F_Y$$

- We discretize the two distributions and now we have the following test :

$$H_0 : \boldsymbol{P}_X = \boldsymbol{P}_Y \quad vs \quad H_1 : \boldsymbol{P}_X \neq \boldsymbol{P}_Y$$

- We can use the following test statistic for the WHom test :

$$T^h_{CWKL} = \frac{2nm}{n+m} D^w_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y)$$

Asymptotics of the T^h_{CWKL} statistics

Theorem (4)

Under the null hypothesis $H_0 : \boldsymbol{P}_X = \boldsymbol{P}_Y$ *we have :*

$$T^h_{CWKL} = \frac{2nmD^W_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y)}{n+m} \xrightarrow[n\to\infty]{\mathcal{D}} \sum_{i=1}^{r} \beta_i^* Z_i^2,$$

where Z_i, $i = 1, \dots, r$ *are iid standard Normal variables,* β_i^*, $i = 1, \dots, r$ *are the eigenvalues of the matrix* $C^* \Sigma_{\hat{\boldsymbol{P}}_X}$ *where* $C^* = (c^*_{ij})_{i,j=1,\dots,K} = \begin{cases} 0, & \text{if } i \neq j \\ \frac{w_i}{\hat{p}_{X,i} \sum_l w_l \hat{p}_{X,l}}, & \text{if } i = j \end{cases},$

$\Sigma_{\hat{\boldsymbol{P}}_X} =$*diag*$(\hat{\boldsymbol{P}}_X) - \hat{\boldsymbol{P}}_X (\hat{\boldsymbol{P}}_X)^T$ *and* $r =$*rank*$(\Sigma_{\hat{\boldsymbol{P}}_X} C^* \Sigma_{\hat{\boldsymbol{P}}_X})$.

Main steps of the proof (1/2)

We can show that the second order Taylor expansion of $D^{W}_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y)$ is :

$$2D^{W}_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y) = (\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y)^T C^* (\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y) + 2o(||\hat{\boldsymbol{P}}_X - \boldsymbol{P}_X||^2) \\ + 2o(||\hat{\boldsymbol{P}}_Y - \boldsymbol{P}_Y||^2),$$

where $\frac{\partial^2 D^{W}_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y)}{\partial \hat{\boldsymbol{P}}_X^2} \overset{H_0}{=}: C^*$. From the Central Limit Theorem we have :

$$\sqrt{n}(\hat{\boldsymbol{P}}_X - \boldsymbol{P}_X) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_{\boldsymbol{P}_X})$$

$$\sqrt{m}(\hat{\boldsymbol{P}}_Y - \boldsymbol{P}_Y) \xrightarrow[m \to \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_{\boldsymbol{P}_Y})$$

and, for $n \to \infty$, $m \to \infty$, we get

$$2D^{W}_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y) \approx (\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y)^T C^* (\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y). \tag{4}$$

Main steps of the proof (2/2)

Under the null hypothesis H_0 we have

$$K = \sqrt{\frac{nm}{m+n}}(\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y) \xrightarrow[n,\ m \to \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_{\boldsymbol{P}_X}) \tag{5}$$

and, from (4) and (5) we obtain :

$$\begin{aligned} & 2\frac{nm}{m+n} D^W_{CKL}(\hat{\boldsymbol{P}}_X, \hat{\boldsymbol{P}}_Y) = \\ & = \sqrt{\frac{nm}{m+n}}(\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y)^T C^* \sqrt{\frac{nm}{m+n}}(\hat{\boldsymbol{P}}_X - \hat{\boldsymbol{P}}_Y) \\ & = K^T C^* K \xrightarrow[n,\ m \to \infty]{\mathcal{D}} \sum_{i=1}^{r} \beta_i^* Z_i^2 . \end{aligned}$$

Plan

Size of WHom

Table – Size of WHom in standard Normal case

standard Normal case			
Partition : $(-\infty,\ -2) \cup (-2,\ 2) \cup (2,\ \infty)$			
n	WHom with $\boldsymbol{w} = (1, 1, 1)^T$	WHom with $\boldsymbol{w} = (2, 1, 2)^T$	WHom with $\boldsymbol{w} = (1, 2, 1)^T$
25	$\simeq 0$	$\simeq$0	$\simeq$0
50	0.0003	0.0003	0.0008
100	0.0098	0.0084	0.0127
200	0.0556	0.0626	0.0533
500	0.0644	0.0612	0.0568
1000	0.0546	0.0589	0.0565

Table – Size of WHom in Log-Normal case

Log-Normal case $(\mu = 0,\ \sigma = 1)$			
Partition : $(0,\ 1) \cup (1,\ 2) \cup (2,\ \infty)$			
n	WHom with $\boldsymbol{w} = (1, 1, 1)^T$	WHom with $\boldsymbol{w} = (2, 1, 2)^T$	WHom with $\boldsymbol{w} = (1, 2, 1)^T$
25	0.0645	0.0703	0.0468
50	0.0578	0.0606	0.0733
100	0.0525	0.0498	0.0647
200	0.0487	0.0547	0.0515
500	0.0472	0.0455	0.0465
1000	0.0499	0.0508	0.0478

Table – Power comparison of WHom test in symmetric case

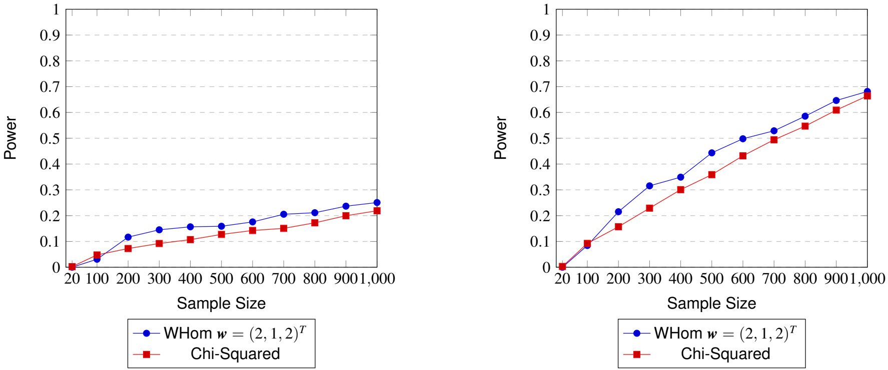
$H_0 : F = F^0 = \mathcal{N}(0, 1)$ vs $H_1 : F = t_p$				
Partition : $(-\infty, -2) \cup (-2, 2) \cup (2, \infty)$				
p	n	WHom with $w = (1, 1, 1)^T$	WHom with $w = (2, 1, 2)^T$	Pearson's Chi-Squared
5	25	0.0001	0.0001	0.0260
	50	0.0335	0.0195	0.1045
	100	0.2581	0.2493	0.2365
	200	0.5782	0.5402	0.4717
	500	0.9236	0.8913	0.8795
	1000	0.9971	0.9959	0.9954
10	25	0.0003	$\simeq 0$	0.0070
	50	0.0078	0.0029	0.0395
	100	0.0884	0.0730	0.0857
	200	0.2276	0.2323	0.1568
	500	0.4400	0.4242	0.3644
	1000	0.6921	0.6813	0.6569
20	25	$\simeq 0$	$\simeq 0$	0.0033
	50	0.0009	0.0010	0.0223
	100	0.0376	0.0404	0.0485
	200	0.1145	0.1208	0.0692
	500	0.1674	0.1586	0.1194
	1000	0.2574	0.2442	0.2217
30	25	$\simeq 0$	$\simeq 0$	0.0032
	50	0.0014	0.0007	0.0197
	100	0.0318	0.0207	0.0400
	200	0.1015	0.0859	0.0569
	500	0.1159	0.1173	0.0853
	1000	0.1470	0.1462	0.1229
$H_0 : F = F^0 = \mathcal{N}(0, 1)$ vs $H_1 : F = Cauchy(0, p)$				
1	25	0.0577	0.1498	0.3816
	50	0.5882	0.6832	0.8980
	100	0.6949	0.9847	0.9977
	200	1	1	1
	500	1	1	1
	1000	1	1	1
0.5	25	0.0040	0.0101	0.0534
	50	0.1255	0.1771	0.3296
	100	0.6068	0.6522	0.6467
	200	0.9362	0.9533	0.9352
	500	1	1	0.9999
	1000	1	1	1

Power of WHom (2/3)

Table – Power comparison of WHom test in non symmetric case

$H_0 : F = F^0 = Log - Normal(0, 1)$ vs $H_1 : F$				
Partition : $(0, 2) \cup (2, 5) \cup (5, \infty)$				
F	n	WHom with $\boldsymbol{w} = (1, 1, 1)^T$	WHom with $\boldsymbol{w} = (1, 2, 5)^T$	Pearson's Chi-Squared
Weibull(1, 1.4)	25	0.0265	0.0021	0.0395
	50	0.0352	0.0247	0.0623
	100	0.0788	0.0635	0.1182
	200	0.1704	0.2136	0.1965
	500	0.4424	0.5474	0.4626
	1000	0.7670	0.8407	0.7850
Gamma(1, 1.4)	25	0.0210	0.0007	0.04154
	50	0.0386	0.0229	0.0588
	100	0.1500	0.1393	0.1130
	200	0.2536	0.3369	0.1998
	500	0.5377	0.5841	0.4704
	1000	0.7989	0.8569	0.7807
IG(3, 1)	25	0.0620	0.0870	0.1550
	50	0.2292	0.2966	0.3209
	100	0.5682	0.5911	0.5998
	200	0.8746	0.9199	0.8961
	500	0.9992	1	0.9997
	1000	1	1	1

Power of WHom (3/3)



(a) WHom test power comparison between $\mathcal{N}(0, 1)$ and t_{20}

(b) WHom test power comparison between $\mathcal{N}(0, 1)$ and t_{10}

Figure – Power comparison of WHom test

Plan

5. Tests for Markov chains

- Time-homogeneous irreducible r-order Markov chain : $(X_\kappa)_{\kappa\in\mathbb{N}}$
- State space : $\mathcal{S} = \{1, 2, \ldots, s\}$
- Transition probability matrix :

$$\boldsymbol{P} = \left(p_{i_1,\ldots,i_r:i_{r+1}} \ : \ i_1,\ldots,i_r \in \mathcal{S}^r, \ i_{r+1} \in \mathcal{S}, \ p_{i_1,\ldots,i_r:i_{r+1}} \geq 0, \ \sum_{i_{r+1}\in\mathcal{S}} p_{i_1,\ldots,i_r:i_{r+1}} = 1 \right)$$

where

$$\mathbb{P}(x_n = i_{r+1} | x_{n-1} = i_r, \ldots, x_{n-r} = i_1) = p_{i_1,\ldots,i_r:i_{r+1}} \quad \forall i_1,\ldots,i_r \in \mathcal{S}^r, \forall i_{r+1} \in \mathcal{S}.$$

- Initial distribution vector : $\boldsymbol{\mu} \in [0,1]^{\mathcal{S}^r}$

$D_\phi^w(\boldsymbol{P},\boldsymbol{Q})$ Divergence

Definition

Weighted ϕ-divergence between Markov chains of general order

Let two r-order Markov chains $(X_\kappa)_{\kappa\in\mathbb{N}}$, $(Y_\kappa)_{\kappa\in\mathbb{N}}$ with state space $\mathcal{S}$ and probability transition matrices $\boldsymbol{P} = (p_{i_1,\dots,i_r:i_{r+1}})_{i_1,\dots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$ and $\boldsymbol{Q} = (q_{i_1,\dots,i_r:i_{r+1}})_{i_1,\dots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$ respectively. Then, the weighted ϕ-divergence between (X_κ) and (Y_κ) is defined as :

$$D_\phi^w(\boldsymbol{P},\boldsymbol{Q}) = \frac{\sum_{i_1,\dots,i_r\in S^r}\sum_{i_{r+1}\in S}\mu^X_{i_1,\dots,i_r} w_{i_1,\dots,i_r:i_{r+1}} q_{i_1,\dots,i_r:i_{r+1}}\,\phi\left(\frac{p_{i_1,\dots,i_r:i_{r+1}}}{q_{i_1,\dots,i_r:i_{r+1}}}\right)}{k},$$

where $\phi \in \Phi =$
$\left\{\phi : \phi \in C^2, \phi(1) = \phi'(1) = 0, \phi''(1) > 0, 0\phi\left(\frac{0}{0}\right) = 0, 0\phi\left(\frac{p}{0}\right) = \lim_{u\to\infty}\frac{\phi(u)}{u}\right\}$
$k = \sum_{i_1,\dots,i_r\in S^r}\sum_{i_{r+1}\in S}\mu^X_{i_1,\dots,i_r} w_{i_1,\dots,i_r:i_{r+1}} p_{i_1,\dots,i_r:i_{r+1}}$ is a constant,
$\boldsymbol{W} = (w_{i_1,\dots,i_r:i_{r+1}})_{i_1,\dots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$ is the influence matrix (with $\boldsymbol{W} \in \mathcal{M}_d(\mathbb{R}_+)$ and $d = dim(\boldsymbol{W}) = dim(\boldsymbol{P}) = dim(\boldsymbol{Q})$).

$D_\phi^w(\boldsymbol{P}, \boldsymbol{Q})$ Divergence

Based on $\boldsymbol{W}$ matrix we can

(i) assume a kind of prior information for each transition (or state),

(ii) direct the ”focus” of the divergence,

(iii) eliminate the impact of useless states, etc.

According to the ϕ function we get

(i) for $\phi(x) = x \log x - x + 1$ the corrected Kullback-Leibler divergence,

(ii) for $\phi(x) = (x-1)^2$ the Pearson’s Chi-Squared divergence, etc.

Homogeneity : Framework

- Assume two r-order Markov chains, $(X_\kappa)_{\kappa\in\mathbb{N}}$ and $(Y_\kappa)_{\kappa\in\mathbb{N}}$, with state space $\mathcal{S}$.

- Unknown transition probability matrices,
$\boldsymbol{P}^x = (p^x_{i_1,\ldots,i_r:i_{r+1}})_{i_1,\ldots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$ and
$\boldsymbol{P}^y = (p^y_{i_1,\ldots,i_r:i_{r+1}})_{i_1,\ldots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$.

- Assume two sequences of observations $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_m)$ raised from the previous models.

- We wish to test the hypothesis :

$$H_0 : \boldsymbol{P}^x = \boldsymbol{P}^y \text{ vs } H_1 : \text{not } H_0, \tag{6}$$

assuming or not prior information on each transition (or state).

- We will use the estimator of the weighted ϕ-divergence $D^w_\phi(\boldsymbol{P}^x, \boldsymbol{P}^y)$.

Homogeneity : Asymptotics

Theorem

Assume the weighted ϕ-divergence $D_\phi^w(\boldsymbol{P}^x, \boldsymbol{P}^y)$ and its estimator $D_\phi^w(\widehat{\boldsymbol{P}}^x, \widehat{\boldsymbol{P}}^y)$. Under the null hypothesis $H_0 : \boldsymbol{P}^x = \boldsymbol{P}^y$ we have :

$$T_\phi^w = \frac{2nm}{\phi''(1)(n+m)} D_\phi^w(\widehat{\boldsymbol{P}}^x, \widehat{\boldsymbol{P}}^y) \xrightarrow[n\to\infty]{\mathcal{D}} \sum_{i=1}^{k} \beta_i Z_i^2,$$

where Z_i, $i = 1, \ldots, k$ are iid $\mathcal{N}(0,1)$ variables, $k =$ rank $(\Sigma_{\boldsymbol{P}^x} C \Sigma_{\boldsymbol{P}^x})$ and β_i, $i = 1, \ldots, k$ are the eigenvalues of the matrix $C\Sigma_{\boldsymbol{P}^x}$ where

$$C_{\mathcal{S}^{r+1}\times\mathcal{S}^{r+1}} = \begin{cases} \mathbf{0}, & \text{if } \boldsymbol{i} \neq \boldsymbol{j} \\ \dfrac{\tilde{w}_{i_1,\ldots,i_r:i_{r+1}}}{p^x_{i_1,\ldots,i_r:i_{r+1}} \sum_{i_1,\ldots,i_r\in\mathcal{S}^r} \sum_{i_{r+1}\in\mathcal{S}} \tilde{w}_{i_1,\ldots,i_r:i_{r+1}} p^x_{i_1,\ldots,i_r:i_{r+1}}}, & \text{if } \boldsymbol{i} = \boldsymbol{j} \end{cases},$$

$\Sigma_{\boldsymbol{P}^x}$ is the covariance matrix and $\tilde{w}_{i_1,\ldots,i_r:i_{r+1}} = \mu^X_{i_1,\ldots,i_r} w_{i_1,\ldots,i_r:i_{r+1}}$.

Goodness-of-fit : Framework

- Assume an r-order Markov chain, $(X_\kappa)_{\kappa\in\mathbb{N}}$, with state space $\mathcal{S}$.
- Unknown transition probability matrix, $\boldsymbol{P}^x = (p^x_{i_1,\dots,i_r:i_{r+1}})_{i_1,\dots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$.
- Assume a sequences of observations $\boldsymbol{x} = (x_1, \dots, x_n)$ raised from the previous model.
- We wish to test the hypothesis :

$$H_0 : \boldsymbol{P}^x = \boldsymbol{P}^0 \text{ vs } H_1 : \text{not } H_0, \tag{7}$$

where $\boldsymbol{P}^0 = (p^0_{i_1,\dots,i_r:i_{r+1}})_{i_1,\dots,i_r\in\mathcal{S}^r, i_{r+1}\in\mathcal{S}}$ is a hypothesized transition probability matrix.

- We will use the estimator of the weighted ϕ-divergence $D^w_\phi(\boldsymbol{P}^x, \boldsymbol{P}^0)$.

Goodness-of-fit : Asymptotics

Theorem

Assume the weighted ϕ-divergence $D_\phi^w(\boldsymbol{P}^x, \boldsymbol{P}^0)$ and its estimator $D_\phi^w(\widehat{\boldsymbol{P}}^x, \boldsymbol{P}^0)$. Under the null hypothesis $H_0 : \boldsymbol{P}^x = \boldsymbol{P}^0$ we have :

$$T_\phi^{*w} = \frac{2n}{\phi''(1)} D_\phi^w(\widehat{\boldsymbol{P}}^x, \boldsymbol{P}^0) \xrightarrow[n\to\infty]{\mathcal{D}} \sum_{i=1}^{k} \beta_i^* Z_i^2,$$

where Z_i, $i = 1, \ldots, k$ are iid $\mathcal{N}(0,1)$ variables, $k =$ rank $(\Sigma_{\boldsymbol{P}^0} C^* \Sigma_{\boldsymbol{P}^0})$ and β_i^*, $i = 1, \ldots, k$ are the eigenvalues of the matrix $C^* \Sigma_{\boldsymbol{P}^0}$ where

$$C^*_{\mathcal{S}^{r+1}\times\mathcal{S}^{r+1}} = \begin{cases} \mathbf{0}, & \text{if } \boldsymbol{i} \neq \boldsymbol{j} \\ \dfrac{\tilde{w}_{i_1,\ldots,i_r:i_{r+1}}}{p^0_{i_1,\ldots,i_r:i_{r+1}} \sum_{i_1,\ldots,i_r\in\mathcal{S}^r} \sum_{i_{r+1}\in\mathcal{S}} \tilde{w}_{i_1,\ldots,i_r:i_{r+1}} p^0_{i_1,\ldots,i_r:i_{r+1}}}, & \text{if } \boldsymbol{i} = \boldsymbol{j} \end{cases},$$

$\Sigma_{\boldsymbol{P}^0}$ is the covariance matrix and $\tilde{w}_{i_1,\ldots,i_r:i_{r+1}} = \mu^X_{i_1,\ldots,i_r} w_{i_1,\ldots,i_r:i_{r+1}}$.

Goodness-of-fit

Remark

The T_ϕ^{*w} test statistic, constitutes a flexible generalization where

1. for a non informative matrix $\boldsymbol{W}$ (e.g. uniform matrix) and $\phi(x) = x \log x$:
$$T_\phi^{*w} = -2 \log \lambda,$$
2. for a non informative matrix $\boldsymbol{W}$ (e.g. uniform matrix) and $\phi(x) = (x-1)^2$:
$$T_\phi^{*w} = X^2.$$

Simulations : homogeneity framework

- Assume two 1-order Markov chains, $(X_\kappa)_{\kappa\in\mathbb{N}}$ and $(Y_\kappa)_{\kappa\in\mathbb{N}}$, with state space $\mathcal{S} = \{1,2,3\}$.
- Transition probability matrices :

$$\boldsymbol{P}^x = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/6 & 1/6 & 3/6 \\ 1/6 & 2/6 & 3/6 \end{pmatrix}, \quad \boldsymbol{P}^y = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/6 & 1/6 & 3/6 \\ \frac{1+6c}{6} & \frac{2+6c}{6} & \frac{3-12c}{6} \end{pmatrix},$$

  with c being a compatibility parameter where $\lim_{c\to 0}(\boldsymbol{P}^y - \boldsymbol{P}^x) = \boldsymbol{0}$.

- We simulate two random sequences from each model, $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_m)$ and we wish to test the hypothesis (1) of homogeneity :

$$H_0 : \boldsymbol{P}^x = \boldsymbol{P}^y \text{ vs } H_1 : \text{not } H_0.$$

Simulations : homogeneity framework

- Test statistics : Pearson's chi-squared X^2, likelihood ratio $-2\log\lambda$, T^{w1}_{CKL}, T^{w2}_{CKL} and T^{w3}_{CKL}. With,

$$\mathbf{W1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{W2} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}, \quad \mathbf{W3} = \begin{pmatrix} |\hat{p}^x_{1:1} - \hat{p}^y_{1:1}| & |\hat{p}^x_{1:2} - \hat{p}^y_{1:2}| & |\hat{p}^x_{1:3} - \hat{p}^y_{1:3}| \\ |\hat{p}^x_{2:1} - \hat{p}^y_{2:1}| & |\hat{p}^x_{2:2} - \hat{p}^y_{2:2}| & |\hat{p}^x_{2:3} - \hat{p}^y_{2:3}| \\ |\hat{p}^x_{3:1} - \hat{p}^y_{3:1}| & |\hat{p}^x_{3:2} - \hat{p}^y_{3:2}| & |\hat{p}^x_{3:3} - \hat{p}^y_{3:3}| \end{pmatrix}.$$

- For the T^w_ϕ test statistics we reject the NULL hypothesis when, $T^w_\phi > q_{1-\alpha}$. Where the critical point verifies the equality $\mathbb{P}\left(\sum_{i=1}^k \beta_i Z_i^2 \leq q_{1-\alpha}\right) = 1-\alpha$.

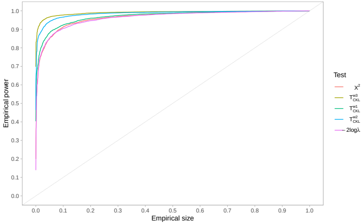
- For the Monte Carlo simulations we use $\alpha = 5\%$ and 10000 repetitions.

Simulations : Power-Size

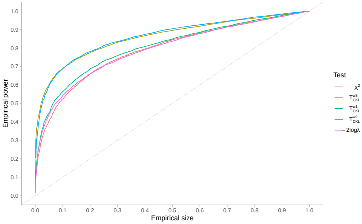
Table – Power-Size comparison of homogeneity test

c	n,m	T^{w1}_{CKL}	T^{w2}_{CKL}	T^{w3}_{CKL}	X^2	$-2\log\lambda$
0	100	0.09283552	0.08173562	0.0881459	0.0532	0.05645992
	500	0.06300000	0.06600000	0.0590000	0.0511	0.05780000
	1000	0.04600000	0.05000000	0.0560000	0.0472	0.05420000
	5000	0.05900000	0.05100000	0.0650000	0.0480	0.05520000
	10000	0.05200000	0.05800000	0.0410000	0.0500	0.05330000
$\frac{1}{10}$	100	0.3038229	0.3722334	0.4395161	0.2439	0.2496421
	500	0.9240000	0.9720000	0.9910000	0.9083	0.9169000
	1000	1	1	1	0.9991	0.9986000
	5000	1	1	1	1	1
	10000	1	1	1	1	1
$\frac{1}{50}$	100	0.1087398	0.120935	0.1093117	0.0624	0.06420793
	500	0.0820000	0.111000	0.0820000	0.0815	0.08410000
	1000	0.1400000	0.168000	0.1410000	0.1205	0.12410000
	5000	0.4940000	0.642000	0.6910000	0.5047	0.50380000
	10000	0.8370000	0.920000	0.9440000	0.8416	0.84770000
$\frac{1}{100}$	100	0.09766022	0.1027467	0.08434959	0.0540	0.05971068
	500	0.06800000	0.0710000	0.060000000	0.0586	0.06110000
	1000	0.07900000	0.0870000	0.06700000	0.0665	0.07140000
	5000	0.14100000	0.1820000	0.18600000	0.1444	0.14110000
	10000	0.25400000	0.3510000	0.35800000	0.2532	0.26540000

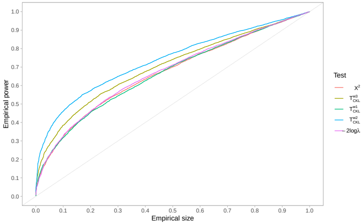
Simulations : ROC analysis



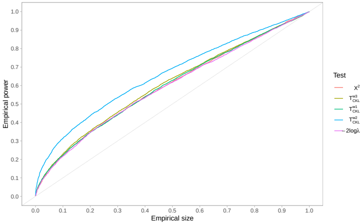
(a) $c = \frac{1}{10}$



(b) $c = \frac{1}{15}$



(c) $c = \frac{1}{20}$



(d) $c = \frac{1}{25}$

Figure – ROC curves for n,m=500.

Plan

5. Conclusion and future work

- WGoF & WHom : Performs better than the most tests for detecting small dissimilarities when we focus in the tail(s) (especially for samples > 1000).
- Accurate and flexible approach. We can test any distribution under the null & We can pay more attention on the subsets we want to focus.
- Next steps : Apply this procedures to inference problems for stochastic processes ; construct an R package
- Already done : Construction of GoF and homogeneity tests for Markov processes.
- In progress : Construction of GoF and homogeneity tests for semi-Markov processes.
- Future work : Relation between the set of the weights and the convergence speed in Theorem 2 (power function).

Plan

6. Bibliography

Barbu, V. S., Karagrigoriou, A., and Preda, V. (2018).
Entropy and divergence rates for Markov chains : II. the weighted case.
Proceedings of the Romanian Academy-series A : Mathematics, Physics, Technical Sciences, Information Science, 19(1) :3–10.

Cressie, N. and Read, T. R. (1984).
Multinomial goodness-of-fit tests.
Journal of the Royal Statistical Society : Series B (Methodological), 46(3) :440–464.

Frank, O., Menéndez, M., and Pardo, L. (1998).
Asymptotic distributions of weighted divergence between discrete distributions.
Communications in Statistics-Theory and Methods, 27(4) :867–885.

Gkelsinis, T. and Barbu, V. S. (2023).
Tests-of-fit for general order Markov chains with asymmetrically important transitions.
submitted.

Gkelsinis, T. and Karagrigoriou, A. (2020).
Theoretical aspects on measures of directed information with simulations.
Mathematics, 8(4) :587.

Gkelsinis, T., Karagrigoriou, A., and Barbu, V. S. (2022).
Statistical inference based on weighted divergence measures with simulations and applications.
Statistical Papers, 63 :1511–1536 ; DOI : https ://doi.org/10.1007/s00362–022–01286–z.

Menéndez, M. L., Pardo, J. A., and Pardo, L. (2001).
Csiszar's ϕ-divergences for testing the order in a Markov chain.
Statistical Papers, 42 :313–328.