THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A,
Volume 19, Number 3/2018, pp. 413–421

MATHEMATICS

ENTROPY AND DIVERGENCE RATES FOR MARKOV CHAINS. III. THE CRESSIE AND READ CASE AND APPLICATIONS

Vlad Stefan BARBU[1], Alex KARAGRIGORIOU[2], Vasile PREDA[3]

[1] Université de Rouen, LMRS, France; e-mail: barbu@univ-rouen.fr
[2] University of the Aegean, Department of Mathematics, Greece; e-mail: alex.karagrigoriou@aegean.gr
[3] University of Bucharest and ISMMA of the Romanian Academy, Romania; e-mail: vasilepreda0@gmail.com
Corresponding author: Vlad Stefan BARBU, Université de Rouen, Laboratoire de Mathématiques Raphaël Salem, UMR 6085, Avenue de l'Université, BP.12, F76801 Saint-Étienne-du-Rouvray, France, e-mail: barbu@univ-rouen.fr

Abstract. In this work we consider generalized Alpha and Beta divergence measure for Markov chains as introduced in [2], where the weighted versions have been investigated in [3]. In continuation to that work, we present generalized Cressie and Read power divergence class of measures, obtain their limiting behavior and numerically investigate some properties of all these generalized divergence measures and rates.

Key words: divergence measures, information measures, Markov chains, entropy, divergence rates, Cressie and Read divergence.

1. PRELIMINARIES

In this section we remind some definitions and results from [2] related to Alpha and Beta divergence measures and provide the basic results on their rates.

Let $(X_n)_{n\in N}$ be an ergodic time-homogeneous Markov chain with finite state space $\chi = \{1,...,M\}.$ For this Markov chain, we consider two different probability laws. Under the first law, let $p_i = P(X_1 = i), i \in \chi,$ denote the initial distribution of the chain and $p_{ij} = P(X_{k+1} = j \mid X_k = i), i, j \in \chi,$ the associated transition probabilities. Let also $\mathbf{p}_n$ denote the joint probability distribution of $(X_1, X_2,..., X_n),$ *i.e.*, $\mathbf{p}_n(i_{1:n}) = p_{i_1} p_{i_1 i_2} ... p_{i_{n-1} i_n}, i_1,..., i_n \in \chi,$ were we denoted by $i_{1:n}$ the n-tuple $(i_1,...,i_n) \in \chi^n$. Similarly we define under the second law q_i, q_{ij}, $\mathbf{q}_n(i_{1:n})$ and $\mathbf{q}_n$. Under this setting of finite state space Markov chains, the Alpha-Gamma measure between the two models is defined as the Alpha-Gamma measure between the two joint probability distributions $\mathbf{p}_n$ and $\mathbf{q}_n$ (cf. [2]), and is written under the normalized form as

$$D_{AG}(\mathbf{p}_n, \mathbf{q}_n) = \frac{1}{\alpha(\alpha - 1)} \log\left(\sum_{i_{1:n} \in \chi^n} \tilde{\mathbf{p}}_n^{\alpha}(i_{1:n}) \tilde{\mathbf{q}}_n^{1-\alpha}(i_{1:n}) \right), \quad (1)$$

where

$$\tilde{\mathbf{p}}_n = \tilde{p}_{i_1} \tilde{p}_{i_1 i_2} \cdots \tilde{p}_{i_{n-1} i_n}, \tilde{\mathbf{q}}_n = \tilde{q}_{i_1} \tilde{q}_{i_1 i_2} \cdots \tilde{q}_{i_{n-1} i_n},$$

with $\tilde{p}_i$, $\tilde{p}_{ij}$, $\tilde{q}_i$ and $\tilde{q}_{ij}$, $i, j \in \chi,$ defined by

$$\tilde{p}_i = \frac{p_i}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{p}_n(i_{1:n})\right)^{1/n}}, \tilde{p}_{ij} = \frac{p_{ij}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{p}_n(i_{1:n})\right)^{1/n}}, \tag{2}$$

$$\tilde{q}_i = \frac{q_i}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{q}_n(i_{1:n})\right)^{1/n}}, \tilde{q}_{ij} = \frac{q_{ij}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{q}_n(i_{1:n})\right)^{1/n}}. \tag{3}$$

Similarly, the Beta-Gamma measure between the two Markov models is defined by (cf. [2])

$$D_{BG}(\mathbf{p}_n,\mathbf{q}_n) = -\frac{1}{\alpha}\log\left(\sum_{i_{1:n}\in\chi^n} \tilde{\mathbf{p}}_n(i_{1:n})\tilde{\mathbf{q}}_n^{\alpha}(i_{1:n})\right), \tag{4}$$

where

$$\tilde{\mathbf{p}}_n = \tilde{p}_{i_1}\tilde{p}_{i_1 i_2}\ldots\tilde{p}_{i_{n-1}i_n}, \tilde{\mathbf{q}}_n = \tilde{q}_{i_1}\tilde{q}_{i_1 i_2}\ldots\tilde{q}_{i_{n-1}i_n},$$

with $\tilde{p}_i$, $\tilde{p}_{ij}$, $\tilde{q}_i$ and $\tilde{q}_{ij}$, $i,j\in\chi$, defined by

$$\tilde{p}_i = \frac{p_i}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{p}_n^{1+\alpha}(i_{1:n})\right)^{1/n(1+\alpha)}}, \tilde{p}_{ij} = \frac{p_{ij}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{p}_n^{1+\alpha}(i_{1:n})\right)^{1/n(1+\alpha)}}, \tag{5}$$

$$\tilde{q}_i = \frac{q_i}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{q}_n^{1+\alpha}(i_{1:n})\right)^{1/n(1+\alpha)}}, \tilde{q}_{ij} = \frac{q_{ij}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{q}_n^{1+\alpha}(i_{1:n})\right)^{1/n(1+\alpha)}}. \tag{6}$$

The following theorems provide the corresponding divergence rates.

THEOREM 1 (cf. [2]). *Under the setting of the present section, we have*

$$\lim_{n\to\infty}\frac{1}{n}D_{BG}(\mathbf{p}_n,\mathbf{q}_n) = -\frac{1}{\alpha}\log\lambda(\alpha),$$

where $\lambda(\alpha) := \lim_{n\to\infty}\lambda_n(\alpha)$ *(assumed to exist), where* $\lambda_n(\alpha)$ *is the largest positive eigenvalue of* $\tilde{R}(n) = (\tilde{r}_{ij}(\alpha))_{i,j\in\chi}$, *where*

$$\tilde{r}_{ij}(\alpha) = \tilde{p}_{ij}\tilde{q}_{ij}^{\alpha} = \frac{p_{ij}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{p}_n^{1+\alpha}(i_{1:n})\right)^{1/n(1+\alpha)}}\left(\frac{q_{ij}}{\left(\sum_{i_{1:n}\in\chi^n} fq_n^{1+\alpha}(i_{1:n})\right)^{1/n(1+\alpha)}}\right)^{\alpha},$$

with $\tilde{p}_{ij}$ *and* $\tilde{q}_{ij}$ *defined in Equations (5) and (6), respectively.*

THEOREM 2 (cf. [2]). *Under the setting of the present section, we have*

$$\lim_{n\to\infty}\frac{1}{n}D_{AG}(\mathbf{p}_n,\mathbf{q}_n) = \frac{1}{\alpha(\alpha-1)}\log\lambda(\alpha),$$

where $\lambda(\alpha)$ *is the largest positive eigenvalue* of $\tilde{R} = (\tilde{r}_{ij}(\alpha))_{i,j\in\chi}$, *where*

$$\tilde{r}_{ij}(\alpha) = \tilde{p}_{ij}^{\alpha}\tilde{q}_{ij}^{1-\alpha} = \frac{p_{ij}^{\alpha}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{p}_n(i_{1:n})\right)^{\alpha/n}}\cdot\frac{q_{ij}^{1-\alpha}}{\left(\sum_{i_{1:n}\in\chi^n} \mathbf{q}_n(i_{1:n})\right)^{(1-\alpha)/n}},$$

with $\tilde{p}_{ij}$ and $\tilde{q}_{ij}$ *defined in Equations (2) and (3), respectively.*

2. CRESSIE AND READ DIVERGENCE RATES FOR MARKOV CHAINS

Let (A,Ω,μ) be a measurable space and μ_p and μ_q some finite measures (not necessarily probability measures) defined on this space, with densities p and q with respect to a certain measure $\mu.$ In this section we are interested in the family of power divergences introduced independently by Cressie and Read (1984) and Liese and Vajda (1987), which is given by

$$I^{CR}\left(p,q\right)=\frac{1}{\alpha(\alpha-1)}\int\left(p^{\alpha}q^{1-\alpha}-q\right)\mathrm{d}\,\mu,\ \ \alpha\in R, \tag{7}$$

where, for $\alpha=0$ and $1,$ it is defined by continuity. The same prolongation by continuity will be used for all the divergence measures considered in the rest of the paper.

Note that the transformation applied to Alpha and Beta divergences (as done in [2]) can also be applied to the Cressie and Read measure given in (7). The resulting measure is given by

$$D_{CRG}(p,q)=\frac{1}{\alpha(\alpha-1)}\log\left(\int p^{\alpha}(x)\tilde{q}^{1-\alpha}(x)\right)\mathrm{d}\,\mu\,(x), \tag{8}$$

which is the normalized Liese and Vajda's measure defined by

$$I^{R,\alpha}(p,q)=\frac{1}{\alpha(\alpha-1)}\log\left(\int p^{\alpha}q^{1-\alpha}d\mu\right),\alpha\neq 0,1, \tag{9}$$

with $\tilde{q}(x)=\frac{q(x)}{\left(\int q(x)d\mu(x)\right)^{1/(1-\alpha)}}$ in place of $q(x)$. In fact, note that CR divergence can be viewed as a special case of D_A divergence.

Let us now introduce the D_{CRG} divergence for Markov chains. This measure, introduced in Equation (8) in the i.i.d. setting, takes the following form in the Markov chain framework:

$$D_{CRG}(\mathbf{p}_n,\mathbf{q}_n)=\frac{1}{\alpha(\alpha-1)}\log\left(\sum_{i_{1:n}\in\chi^n}\mathbf{p}_n^{\alpha}(i_{1:n})\frac{\mathbf{q}_n^{1-\alpha}(i_{1:n})}{\left(\sum_{i_{1:n}\in\chi^n}\mathbf{q}_n(i_{1:n})\right)^{1-\alpha}}\right). \tag{10}$$

This can be written under the normalized form

$$D_{CRG}(\mathbf{p}_n,\mathbf{q}_n)=\frac{1}{\alpha(\alpha-1)}\log\left(\sum_{i_{1:n}\in\chi^n}\mathbf{p}_n^{\alpha}(i_{1:n})\tilde{\mathbf{q}}_n^{1-\alpha}(i_{1:n})\right), \tag{11}$$

where $\tilde{\mathbf{q}}_n=\tilde{q}_{i_1}\tilde{q}_{i_1i_2}...\tilde{q}_{i_{n-1}i_n},$ with $\tilde{q}_i$ and $\tilde{q}_{ij},$ $i,j\in\chi,$ defined by

$$\tilde{q}_i=\frac{q_i}{\left(\sum_{i_{1:n}\in\chi^n}\mathbf{q}_n(i_{1:n})\right)^{1/n(1-\alpha)}},\tilde{q}_{ij}=\frac{q_{ij}}{\left(\sum_{i_{1:n}\in\chi^n}\mathbf{q}_n(i_{1:n})\right)^{1/n(1-\alpha)}}. \tag{12}$$

The following result concerns the limiting behavior of D_{CRG}.

THEOREM 3. *Under the setting presented before, we have*

$$\lim_{n\to\infty}\frac{1}{n}D_{CRG}(\mathbf{p}_n,\mathbf{q}_n)=\frac{1}{\alpha(\alpha-1)}\log\lambda(\alpha),$$

where $\lambda(\alpha)$ is the largest positive eigenvalue of $\tilde{R}=(\tilde{r}_{ij}(\alpha))$, with

$$\tilde{r}_{ij}(\alpha)=p_{ij}^{\alpha}\tilde{q}_{ij}^{1-\alpha}=\frac{p_{ij}^{\alpha}q_{ij}^{1-\alpha}}{\left(\sum_{i_{1:n}\in\chi^n}\mathbf{q}_n(i_{1:n})\right)^{1/n}}$$

and $\tilde{q}_{ij}$ defined in (12).

3. NUMERICAL APPLICATIONS

In this section we will consider numerical examples in order to illustrate the results obtained in the previous section.

Let $(X_n)_{n\in N}$ be a time-homogeneous two-state Markov chain. As previously described, for this Markov chain we consider two different probability laws, the first one given by a Markov transition matrix $\mathbf{p}=(p_{i,j})_{i,j=1,2}$ and an initial distribution $p=\begin{pmatrix}p_1 & p_2\end{pmatrix}$, while the second one is governed by a Markov transition matrix $\mathbf{q}=(q_{i,j})_{i,j=1,2}$ and an initial distribution $q=\begin{pmatrix}q_1 & q_2\end{pmatrix}$. We consider the transition matrices given by

$$\mathbf{p}=\begin{pmatrix}0.9 & 0.1\\ 0.6 & 0.4\end{pmatrix}\text{and }\mathbf{q}=\begin{pmatrix}0.2 & 0.8\\ 0.1 & 0.9\end{pmatrix},$$

while for the initial distributions we take the corresponding stationary ones, namely

$$\begin{pmatrix}p_1 & p_2\end{pmatrix}=\begin{pmatrix}6/7 & 1/7\end{pmatrix}\text{and}\begin{pmatrix}q_1 & q_2\end{pmatrix}=\begin{pmatrix}1/9 & 8/9\end{pmatrix}.$$

First, the results of Theorem 1, concerning the divergence rate of Beta-Gamma measure, are illustrated in Table 1.

Table 1

The rate of Beta-Gamma divergence

	***n* = 10**	***n* = 15**	***n* = 20**	**rate**
α = 1	BG/*n* = 0.8206	BG/*n* = 0.7990	BG/*n* = 0.7882	0.7533
α = 0.5	BG/*n* = 1.0845	BG/*n* = 1.0707	BG/*n* = 1.0638	1.0403
α = 0. 1	BG/*n* = 1.1391	BG/*n* = 1.1263	BG/*n* = 1.1199	1.0985
α = 0.01	BG/*n* = 1.1306	BG/*n* = 1.1173	BG/*n* = 1.1107	1.0889
α = 0.001	BG/*n* = 1.1295	BG/*n* = 1.1161	BG/*n* = 1.1094	1.0876
KL	KL/*n* = 1.1293	KL/*n* = 1.1160	KL/*n* = 1.1093	1.0892

Second, in Table 2 we illustrate the convergence of Alpha-Gamma measure to the KL measure, as α goes to 1 (cf. Remark 3). Note that the results in Tables 1 and 2 demonstrate both the convergence of the appropriate measure to the corresponding rate as well as the convergence to KL for any value of n (including the limit).

Divergence measures, like the ones discussed in this work, are used as indices of similarity or dissimilarity between populations. As a result, they can be used as a way to evaluate the distance (divergence) between any two populations or functions. Measures of divergence can be used in statistical inference for estimating purposes (Toma [9] and [10]), in the construction of test statistics for tests of fit (*e.g.* Zografos *et al.* [12], Huber-Carol *et al.* [7] and Zhang [11]) or in statistical modeling for the construction of model selection criteria like the Kullback-Leibler measure which has been used for the development of various criteria (*e.g.* Akaike [1] and Cavanaugh [4]).

Table 2

Convergence of Alpha-Gamma measure to the KL measure, as $\alpha \to 1$

	***n* = 10**	***n* = 15**	***n* = 20**	**rate**
α = –1	AG/*n* = 0.4140	AG/*n* = 0.3979	AG/*n* = 0.3898	0.3655
α = 0.5	AG/*n* = 1.0198	AG/*n* = 0.9947	AG/*n* = 0.9822	0.9447
α = 0. 9	AG/*n* = 1.1501	AG/*n* = 1.1352	AG/*n* = 1.1278	1.1055
α = 0.95	AG/*n* = 1.1419	AG/*n* = 1.1279	AG/*n* = 1.1209	1.0998
α = 0.99	AG/*n* = 1.1321	AG/*n* = 1.1187	AG/*n* = 1.1119	1.0917
KL	KL/*n* = 1.1293	KL/*n* = 1.1160	KL/*n* = 1.1093	1.0892

One of the most popular statistics is the Cressie and Read power divergence statistics (CR). The rate of the generalized form of this divergence for Markov sources was derived in Theorem 3. The CR family of statistics was originally proposed for testing the fit of observed frequencies to expected frequencies. Through this family of statistics, Cressie and Read succeeded in providing a unified approach to goodness-of-fit testing for multinomial models. The importance of the proposed statistics lies on the fact that several goodness-of-fit tests can be reduced to test a null hypothesis from a multinomial population and therefore a statistic that measures how much two distributions differ is of high importance. Several well-known test statistics are members of the Cressie and Read family of divergences like the Pearson's chi-square, the likelihood disparity (generating the log-likelihood ratio statistic), the (twice and squared) Hellinger distance (Freeman and Tukey [6]), the Kullback-Leibler divergence and the Neyman modified chi-square which are indexed by $\alpha = 2$, 1 (by continuity), 1/2, 0 (by continuity) and –1, respectively.

In reference to Theorem 3, some results related to CR family of measures are presented below. More precisely, for the Cressie and Read Generalized measure we present the convergence of the measure and associated rate to the appropriate KL measure and rate, as α goes to 0, $\alpha > 0$, and α goes to 1, $\alpha < 1$ (cf. Table 3).

Note that

$$\lim_{a\to 0}\frac{1}{n}D_{CRG}(\mathbf{p}_n,\mathbf{q}_n) = D_{KL}(\mathbf{q}_n,\mathbf{p}_n),$$

while

$$\lim_{a\to 1}\frac{1}{n}D_{CRG}(\mathbf{p}_n,\mathbf{q}_n) = D_{KL}(\mathbf{p}_n,\mathbf{q}_n)$$

which is clearly confirmed by the results in Table 3.

In Table 4 we illustrate the rates of three particular cases of the Generalized Cressie and Read measure: Pearson's χ^2 ($\alpha = 2$), Freeman-Tukey's F^2 ($\alpha = 0.5$) and Neyman χ^2 (Euclidian log-likelihood ratio statistic) ($\alpha = -1$). We have also included the special case $\alpha = -2/3$. Note that, based on a comparative study, this special value was recommended by Read and Cressie [8] (a value between the Pearson's chi-square and the Neyman's chi-square statistic) as a compromise candidate among the different test statistics, although they noted several desirable properties of the other test statistics, including the Pearson's chi-square (see, *e.g.* Section 4.5, Section 6.7, and Appendix A11 of Read and Cressie [8]).

Table 3

Convergence of Cressie and Read Generalized measure/rate to the KL measure/rate, as $a \to 0, a > 0$ and $a \to 1, a < 1$

	***n* = 10**	***n* = 15**	***n* = 20**	**rate**
α = 0.1	CRG/*n* = 0.7682	CRG/*n* = 0.7421	CRG/*n* = 0.7290	0.6899
α = 0.01	CRG/*n* = 0.7215	CRG/*n* = 0.6961	CRG/*n* = 0.6835	0.6455
α = 0.001	CRG/*n* = 0.7170	CRG/*n* = 0.6918	CRG/*n* = 0.6792	0.6413
KL(q,p)	KL/*n* = 0.7166	KL/*n* = 0.6913	KL/*n* = 0.6787	0.6408
α = 0.9	CRG/*n* = 1.1501	CRG/*n* = 1.1352	CRG/*n* = 1.1278	1.1055
α = 0.99	CRG/*n* = 1.1321	CRG/*n* = 1.1187	CRG/*n* = 1.1119	1.0917
α = 0.999	CRG/*n* = 1.1296	CRG/*n* = 1.1162	CRG/*n* = 1.1096	1.0895
KL(p,q)	KL/*n* = 1.1293	KL/*n* = 1.1160	KL/*n* = 1.1093	1.0892

Table 4

The rate of the Generalized Cressie and Read measure for some important special cases: Pearson's $\chi^2(a=2)$, Freeman-Tukey's $F^2(a=0.5)$, Cressie and Read $(a=-2/3)$ and Neyman $\chi^2(a=-1)$

	***n* = 10**	***n* = 15**	***n* = 20**	**rate**
α = 2	CRG/*n* = 0.7253	CRG/*n* = 0.7171	CRG/*n* = 0.7130	0.7007
α = 0.5	CRG/*n* = 1.0198	CRG/*n* = 0.9947	CRG/*n* = 0.9822	0.9447
α = –2/3	CRG/*n* = 0.4839	CRG/*n* = 0.4652	CRG/*n* = 0.4558	0.4277
α = –1	KL/*n* = 0.4140	KL/*n* = 0.3979	KL/*n* = 0.3898	0.3655

In Fig. 1 the Generalized Cressie and Read divergence rate is illustrated as a function of α for the two probability laws of the Markov chain considered at the beginning of this section. We also represented the value of $\frac{1}{n}D_{CRG}$ for several values of *n*. Notice the fast convergence of $\frac{1}{n}D_{CRG}$ to the rate according to Theorem 3. Notice further that, even for small values of *n*, $\frac{1}{n}D_{CRG}$ gives a good approximation of the rate.
In reference to the special value of $\alpha = -2/3$, we observe in Fig. 1 that this is not the value of the index α that discriminates the most between the two Markov chains. For this particular example the value that maximizes the divergence rate is $\alpha^* = 0.88$. Although α^* may be of limiting significance if the two sources are well separated, it will be of great importance in case the two sources are close to each other. Indeed, consider the following example, for which the divergence rate will be expected to be close to 0. Let a time-homogeneous two-state Markov chain $(X_n)_{n\in N}$ evolve under two different probability laws than those of the beginning of this section, the first one given by a Markov transition matrix $\mathbf{p} = (p_{i,j})_{i,j=1,2}$ and an initial distribution $p = \begin{pmatrix} p_1 & p_2 \end{pmatrix}$, while the second one is governed by a Markov transition matrix $\mathbf{q} = (q_{i,j})_{i,j=1,2}$ and an initial distribution $q = \begin{pmatrix} q_1 & q_2 \end{pmatrix}$. We consider the transition matrices given by

$$\mathbf{p} = \begin{pmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{pmatrix} \text{and } \mathbf{q} = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix},$$

while for the initial distributions we take the corresponding stationary ones.

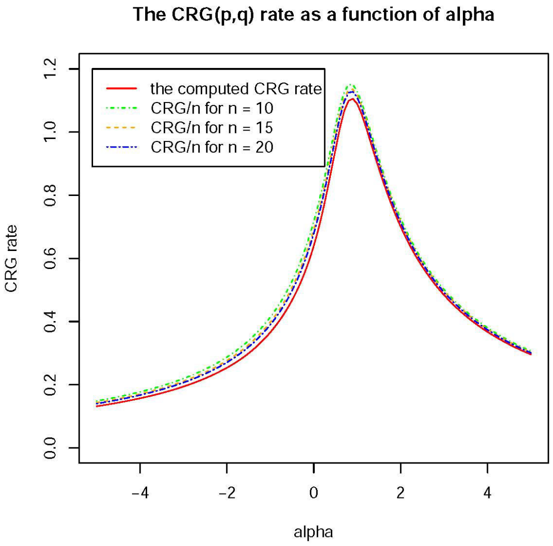

Fig. 1 – The convergence of $\frac{1}{n}D_{CRG}$ w.r.t. n.

Figure 2 presents the Generalized Cressie and Read divergence as a function of α for this example. Note that Figure 2 confirms the closeness of the sources but at the same time provides the value α^* of the index for which the rate is maximized ($\alpha^* = -1.57$). In conclusion, for discriminatory purposes and consequently for statistical inference (*i.e.*, goodness of fit tests, model selection, etc.) we recommend the use of the divergence rate with the index taken to be equal to the value α^*. Note that the same recommendation applies not only to the rate but also to the divergence itself.

Let us now consider two additional examples of two different probability laws governing a Markov chain. First, we are interested in a two-state Markov chain and we set the Markov transition matrices **p** and **q**

$$\mathbf{p} = \begin{pmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{pmatrix} \text{ and } \mathbf{q} = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix},$$

while the initial distribution $p = (p_1 \quad p_2)$ and $q = (q_1 \quad q_2)$ are taken to be the associated stationary ones.

In Fig. 3 we present both the Generalized Cressie and Read divergence rate computed for (p_n, q_n) and also for (q_n, p_n) as a function of α.

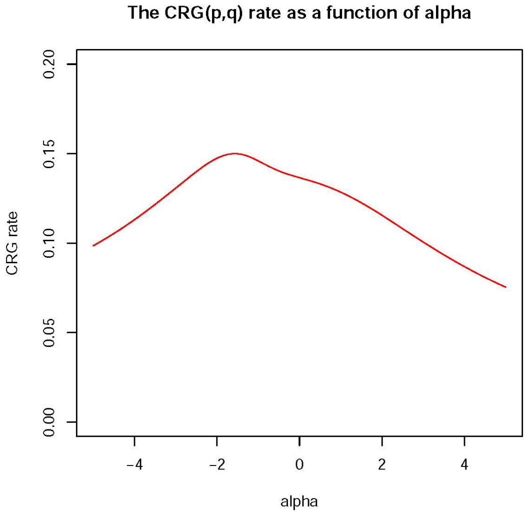

Fig. 2 – The Generalized Cressie and Read divergence rate for the second example.

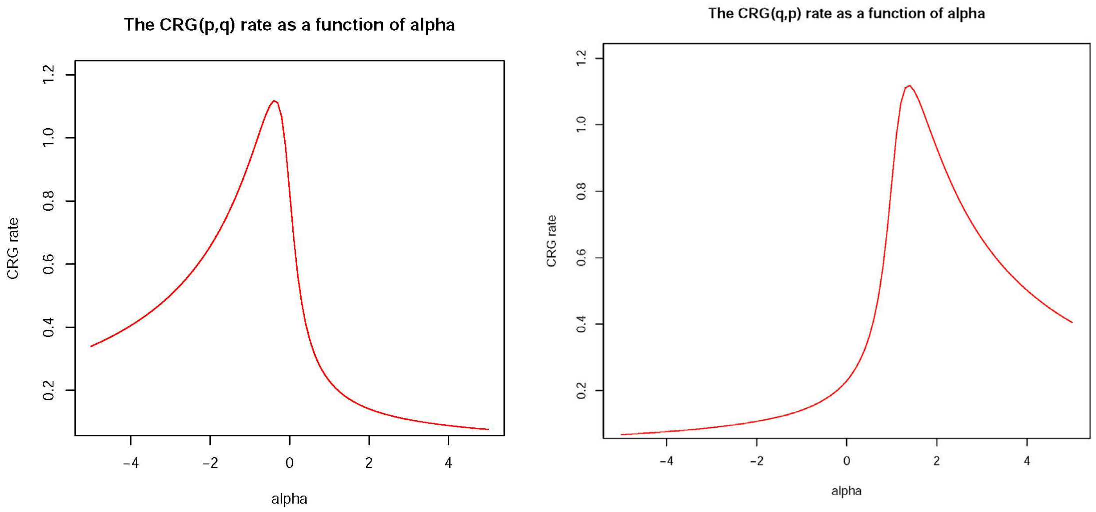

Fig. 3 – Reflection property of the GCR divergence rate for the third example.

Second, we consider another example of two laws governing now a three-state Markov chain. Let **p** and **q** be two Markov transition matrices given by

$$\mathbf{p}=\begin{pmatrix}0.7 & 0.2 & 0.1\\ 0.4 & 0.5 & 0.1\\ 0.4 & 0.1 & 0.5\end{pmatrix} \text{and } \mathbf{q}=\begin{pmatrix}0.9 & 0.05 & 0.05\\ 0.6 & 0.2 & 0.2\\ 0.1 & 0.2 & 0.7\end{pmatrix},$$

while the initial distribution $p=\begin{pmatrix}p_1 & p_2\end{pmatrix}$ and $q=\begin{pmatrix}q_1 & q_2\end{pmatrix}$ are the associated stationary ones.

As for the previous example, in Fig. 4 we present both the Generalized Cressie and Read divergence rate computed for (p_n,q_n) and also for (q_n,p_n) as a function of α.

Note that in both Figs. 3 and 4 there is a symmetry between the two graphs. In fact this phenomenon is due to a reflection property of the GCR divergence. More precisely, let us denote by $D_{CRG;\alpha}(p_n,q_n)$ the GCR divergence evaluated at α. Then, one can easily verify that $D_{CRG;\alpha}(p_n,q_n)=D_{CRG;1-\alpha}(q_n,p_n)$. Obviously, due to Theorem 3 this property holds true also for the divergence rate. For this reason, in Figures 3 and 4 we have a reflection wrt the line $x=0.5$.

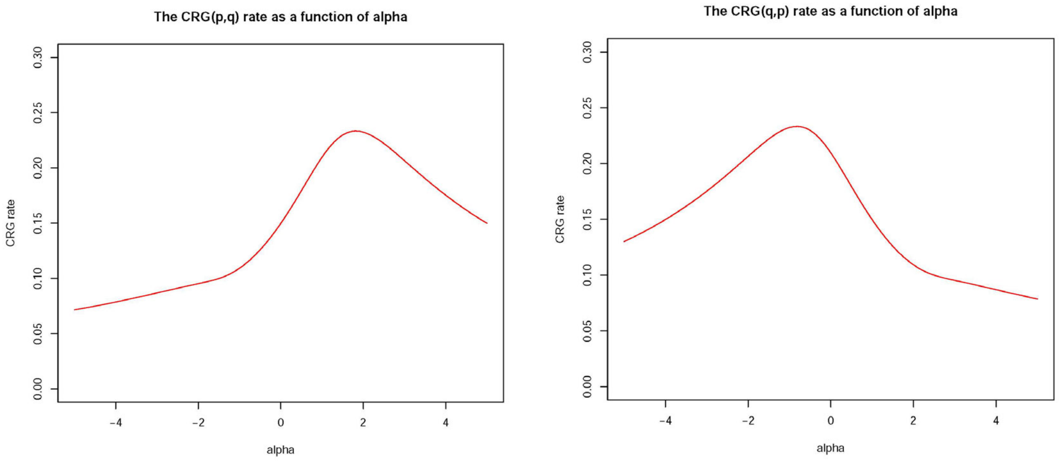

Fig. 4 – Reflection property of the GCR divergence rate for the three-state example.

ACKNOWLEDGEMENTS

The authors would like to thank their colleague Nicolas Vergne from Laboratoire de Mathématiques Raphaël Salem, University of Rouen, France, for his suggestions and help on some technical problems related to this paper. The first two authors would also like to express their appreciation to the University of Rouen and University of Aegean for the opportunity to exchange several visits in both institutions. The research work of Vlad Stefan Barbu was partially supported by the projects *XTerM–Complex Systems, Territorial Intelligence and Mobility* (2014–2018) and *MOUSTIC–Random Models and Statistical, Informatics and Combinatorics Tools* (2016–2019), within the Large Scale Research Networks from the Region of Normandy, France.

REFERENCES

1. H. AKAIKE, *Information theory and an extension of the maximum likelihood principle*, Proceeding of the Second International Symposium on Information Theory, B.N. Petrov and F. Csaki (eds.), Akademia Kaido, Budapest, 1973, pp. 267–281.
2. V.S. BARBU, A. KARAGRIGORIOU, V. PREDA, *Entropy and divergence rates for Markov chains: I. The Alpha-Beta and Alpha-Gamma case*, submitted, 2017.
3. V.S. BARBU, A. KARAGRIGORIOU, V. PREDA, *Entropy and divergence rates for Markov chains: II. The weighted case*, submitted, 2017.
4. J.E. CAVANAUGH, *Criteria for linear model selection based on Kullback's symmetric divergence*, Australian and New Zealand Journal of Statistics, **46**, pp. 257–274, 2004.
5. N. CRESSIE, T.R.C. READ, *Multinomial goodness-of-fit tests*, J. R. Statist. Soc, **5**, pp. 440–454, 1984.
6. M.F. FREEMAN, J.W. TUKEY, *Transformations related to the angular and the square-root*, Ann. Math. Statist., **21**, pp. 607–611, 1950.
7. C. HUBER-CAROL, N. BALAKRISHNAN, M.S. NIKULIN, M. MESBAH, *Goodness-of-fit Tests and Model Validity*, Birkhäuser, Boston, 2002.
8. T.R.C. READ, N. CRESSIE, *Goodness-of-Fit Statistics for Discrete Multivariate Data*, New York, Springer-Verlag, 1988.
9. A. TOMA, *Minimum Hellinger distance estimators for multivariate distributions from the Johnson system*, J. Statist. Plan. and Infer., **138**, pp. 803–816, 2008.
10. A. TOMA, *Optimal robust M-estimators using divergences*, Statistics and Prob. Letters, **79**, pp. 1–5, 2009.
11. J. ZHANG, *Powerful goodness-of-fit tests based on likelihood ratio*, J. R. Stat. Soc. Ser. B, **64**, *2*, pp. 281–294, 2002.
12. K. ZOGRAFOS, K. FERENTINOS, T. PAPAIOANNOU, Φ *-divergence statistics: Sampling properties, multinomial goodness of fit and divergence tests*, Comm. in Statist. Theor. Meth., **19**, *5*, pp. 1785–1802, 1990.

Received March 31, 2017