Advanced Studies in Contemporary Mathematics
32 (2022), No. 3, pp. 347 - 359

www.jangjeon.or.kr
http://dx.doi.org/10.17777/ascm2022.32.3.347

On the Metric Topology Induced from the BIC-based Consistent Metric between Markovian Processes

James R. Bozeman*

1 Abstract

In recent papers (see reference list) the authors Garcia, Gholizadeh and Gonzalez-Lopez develop a local metric from a distance measure between samples coming from discrete Markovian processes. This metric is associated to the Bayesian Information Criterion (BIC). They use the metric to decide if 2 independent samples are governed by the same Stochastic Law. Their results are applied to many topics, including internet navigation patterns, lines of production and Zika strains, for example.

In this paper we expand on the measure above by examining the metric (topological) space induced by it. For instance, we exhibit the open sets in this topology, which are of course the unions of open balls under the metric. We then look at the bases for this topology and check for other topological properties. We show that the metric space cannot be compact and later form an identification space. We conclude by interpreting the work of the authors above in the topological context as well as providing other applications and results, e.g. to other vector-borne diseases and to the DNA molecule..

Keywords: Bayesian Information Criterion, Markov Processes, Metric Space, Stochastic Processes, Topological Space, Vector-borne diseases, Z-DNA

2 Introduction

In [**2**] the authors Jesus E. Garcia, R. Gholizadeh, and Veronica A. Gonzalez-Lopez develop a BIC-based consistent metric for Markovian processes (see the next section for definitions). They show that the metric is statistically consistent and in the case where the stochastic

*Corresponding author, The American University of Malta, Bormla, Malta, james.bozeman@aum.edu.mt

laws are not the same use the metric to find where the discrepancies are. In this paper their results are applied to lines of production.

In [**1**] the authors Jesus E. Garcia and Veronica A. Gonzalez-Lopez show that the Bayesian Information Criterion (BIC) can be used to obtain a consistent estimation of the partition of a Markov process. This partition comes from an equivalence relation based on the transition probabilities of elements of the state space. The authors then go on to introduce a measure to quantify the distance between the parts of a partition. This measure also forms a metric (see the next section for details and proofs). The authors use their results to model internet navigation patterns. In further work these three authors, along with others, have considered a large variety of real-world problems, including the statistical profile of strains of zika (see reference list).

In this paper we expand on the distance measures noted above by considering the topology induced by the metric. This topological (metric) space is studied. In the first case we show that the induced metric (topological) space cannot be compact. In the second case the identification space is formed from the equivalence classes. Then more applications are introduced, in particular to vector-borne diseases and to the left-handed version of the DNA molecule, Z-DNA.

This paper is organized as follows: Section 3 provides the preliminaries, in particular the proofs of the equivalence relation and metric noted above. Section 4 defines the metric space in the context of topological spaces and the results noted above are shown. Section 5 provides the statistical interpretation. In Section 6 we exhibit applications, especially with regard to vector-borne diseases and Z-DNA. We conclude in Section 7.

3 Preliminaries

We closely follow [2], where the authors endeavour to decide if 2 samples of a Markov process follow the same stochastic law. They use the perspective of the Bayesian Information Criterion (BIC) of Schwarz [3]. This is shown to be consistent by the authors in [1]:

Let $(X_t), t\varepsilon\mathbf{N}$, be a discrete time Markov chain, with a finite alphabet A and finite order M. Then let $S = A^M$ be the state space. Denote the string $a_m a_{m+1}...a_n$ by a_m^n, where $a_i \varepsilon A, m \leq i \leq n$. For each $a\varepsilon A$ and $s\varepsilon S$ let $P(a|s) = Prob(X_t = a|X_{t-a}^{t-1} = s)$.

Let x_1^n be a sample of the process (X_t) where $s\varepsilon S$, $a\varepsilon A$. First, $N_n(s)$ is the number of occurrences of the string s in the sample x_1^n. Next $N_n(s,a)$ is the number of occurrences of the string s followed by a in the sample x_1^n. Then if $N_n(s) \neq 0$ the estimator of $P(a|s)$ is $\frac{N_n(s,a)}{N_n(s)}$. The likelihood of the sample $P(X_t = x_1^n)$ is $P(x_1^M)\Pi_{a\varepsilon A, r\varepsilon S} P(a|r)^{N_n(r,a)}$. Then the total number of parameters to be estimated is $(|A|-1)|S|$ which is the cardinal of the set $P(a|s), s\varepsilon S, a\varepsilon A$ where $||$ means cardinality.

The BIC is then defined through the *modified maximum likelihood* ML (see Csiszar and Shields [4]).

$$ML(S, x_1^n) = \Sigma_{a\varepsilon A, r\varepsilon S} N_n(r,a) ln \frac{N_n(r,a)}{N_n(r)}, N_n(r) \neq 0, \forall r \varepsilon S$$

Definition 1: Let (X_t) be a discrete time order M Markov chain, with a finite alphabet A and finite order M. Let $S = A^M$ be the state space where x_1^n is a sample of the process. Then

$$BIC(x_1^n) = ML(S, x_1^n)) - \frac{(|A|-1)|}{\alpha}|S| ln(n)$$

where $\alpha > 0, \varepsilon \mathbf{R}$.

Now consider two independent Markov chains $(X_{1,t})$ and $(X_{2,t})$, $t\varepsilon\mathbf{N}$ of order M with finite alphabet A, state space $S = A^M$ and independent samples $x_{1,1}^{n_1}$ and $x_{2,1}^{n_2}$, respectively. Then let $P(a|s)$ on $(X_{1,t})$ and $Q(a|s)$ on $(X_{2,t})$ be the set of conditional probabilities, respectively, over all $a\varepsilon A$. Now let $N_{n_1+n_2}(s,a) = N_{n_1}(s,a) + N_{n_2}(s,a)$ and $N_{n_1+n_2}(s) = N_{n_1}(s) + N_{n_2}(s)$, where these are as above, computed on the samples $x_{1,1}^{n_1}$ and $x_{2,1}^{n_2}$, respectively. Now to get the BIC for the joint model, with $P(\cdot|r) \neq Q(\cdot|r), \forall r\varepsilon S$, and the 2 samples independent, the likelihood of the 2 samples is

$$P(x_{1,1}^M) P(x_{1,2}^M) \Pi_{a\varepsilon A, r\varepsilon S} P(a|r)^{N_{n_1}(r,a)} Q(a|r)^{N_{n_2}(r,a)}$$

and the total number of parameters to be estimated is now $2(|A|-1)|S|$. Then the log-maximum likelihood for the 2 samples is

$$\Sigma_{a\varepsilon A, r\varepsilon S}\{N_{n_1}(r,a) ln \frac{N_{n_1}(r,a)}{N_{n_1}(r)} + N_{n_2}(r,a) ln \frac{N_{n_2}(r,a)}{N_{n_2}(r)}\}$$

.

In the case where $P(\cdot|s) = Q(\cdot|s)$ for a specific $s\varepsilon S$, the number of parameters are $(|A| - 1)(2|S| - 1)$ and the formula becomes

$$\Sigma_{a\varepsilon A, r\varepsilon(S-s)}\{N_{n_1}(r,a) ln \frac{N_{n_1}(r,a)}{N_{n_1}(r)} + N_{n_2}(r,a) ln \frac{N_{n_2}(r,a)}{N_{n_2}(r)}\} + \Sigma_{a\varepsilon A}(N_{n_1+n_2}(s,a) ln(\frac{N_{n_1+n_2}(s,a)}{N_{n_1+n_2}(s)}))$$

We can now define the metric to decide how far or near the processes are and relate it to the BIC:

Definition 2: Consider two Markov chains $(X_{1,t})$ and $(X_{2,t})$, $t\varepsilon\mathbf{N}$ of order M with finite alphabet A, state space $S = A^M$ and independent samples $x_{1,1}^{n_1}$ and $x_{2,1}^{n_2}$, respectively. Define for a string $s\varepsilon S$, $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) =$

$$\frac{\alpha}{(|A|-1) ln(n_1+n_2)} \sum_{a\varepsilon A}\{N_{n_1}(s,a) ln \frac{N_{n_1}(s,a)}{N_{n_1}(s)} + N_{n_2}(s,a) ln \frac{N_{n_2}(s,a)}{N_{n_2}(s)} - N_{n_1+n_2}(s,a) ln \frac{N_{n_1+n_2}(s,a)}{N_{n_1+n_2}(s)})\}$$

Theorem 3: Consider 3 Markov chains $(X_{1,t})$, $(X_{2,t})$ and $(X_{3,t})$, $t\varepsilon\mathbf{N}$ of order M with finite alphabet A, state space $S = A^M$ and independent samples $x_{1,1}^{n_1}$, $x_{2,1}^{n_2}$ and $x_{3,1}^{n_3}$, respectively. Let $s\varepsilon S$ be a string such that $N_{n_i}(s) \neq 0$, for $i = 1, 2, 3$. Then,

1.) $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) \geq 0$ with equality iff $\frac{N_{n_1}(s,a)}{N_{n_1}(s)} = \frac{N_{n_2}(s,a)}{N_{n_2}(s)} \forall a\varepsilon A$;

2.) $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) = d_s(x_{2,1}^{n_2}, x_{1,1}^{n_1})$;

3.) $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) \leq d_s(x_{1,1}^{n_1}, x_{3,1}^{n_3}) + d_s(x_{3,1}^{n_3}, x_{2,1}^{n_2})$.

Proof 3: The proofs of 1.) and 3.) are given in **[2]** . We include the proof of 2.) for completeness: Noting that $N_{n_1+n_2}(s,a) = N_{n_1}(s,a) + N_{n_2}(s,a) = N_{n_2}(s,a) + N_{n_1}(s,a) = N_{n_2+n_1}(s,a)$ and $N_{n_1+n_2}(s) = N_{n_1}(s) + N_{n_2}(s) = N_{n_2}(s) + N_{n_1}(s) = N_{n_2+n_1}(s)$, we have $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) =$

$$\frac{\alpha}{(|A|-1)ln(n_1+n_2)} \sum_{a\varepsilon A} \{N_{n_1}(s,a) ln \frac{N_{n_1}(s,a)}{N_{n_1}(s)} + N_{n_2}(s,a) ln \frac{N_{n_2}(s,a)}{N_{n_2}(s)} - N_{n_1+n_2}(s,a) ln \frac{N_{n_1+n_2}(s,a)}{N_{n_1+n_2}(s)})\}$$

$=$

$$\frac{\alpha}{(|A|-1)ln(n_2+n_1)} \sum_{a\varepsilon A} \{N_{n_2}(s,a) ln \frac{N_{n_2}(s,a)}{N_{n_2}(s)} + N_{n_1}(s,a) ln \frac{N_{n_1}(s,a)}{N_{n_1}(s)} - N_{n_2+n_1}(s,a) ln \frac{N_{n_2+n_1}(s,a)}{N_{n_2+n_1}(s)})\}$$

$= d_s(x_{2,1}^{n_2}, x_{1,1}^{n_1}) =$ and the proof is complete •.

Theorem 4: Consider two Markov chains $(X_{1,t})$ and $(X_{2,t})$, $t\varepsilon\mathbf{N}$ of order M with finite alphabet A, state space $S = A^M$ and independent samples $x_{1,1}^{n_1}$ and $x_{2,1}^{n_2}$, respectively. Then let $\{P(a|r)\}_{a\varepsilon A, s\varepsilon s}$ on $(X_{1,t})$ and $\{Q(a|r)\}_{a\varepsilon A, s\varepsilon s}$ on $(X_{2,t})$ be the set of conditional probabilities, respectively. Let $s\varepsilon S$ be a string such that $N_{n_i}(s) \neq 0$, for $i = 1, 2$. Then

$$BIC(x_{1,1}^{n_1}, x_{2,1}^{n_2}) < BIC(x_{1,1}^{n_1}, x_{2,1}^{n_2}, =_s) \iff d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) < 1$$

where the second BIC is calculated assuming $P(a|s) = Q(a|s), \forall a\varepsilon A$.

Proof 4: See the proof of Theorem 2 in [2]. •

This Theorem means that if the value of $d_s < 1$ then the BIC indicates the samples as being governed by the same stochastic law. Hence if there is an $a\varepsilon A$ such that $d_s > 1$ then we can say the samples are governed by different stochastic laws. The authors and colleagues use this interpretation on many applications (see reference list).

We now state a Theorem to be used in the sequel on Metric Spaces and which shows the statistical consistency of the metric. It also indicates that as sample sizes increase that d_s is better able to detect differences or similarities between stochastic laws.

Theorem 5: Consider two Markov chains $(X_{1,t})$ and $(X_{2,t})$, $t\varepsilon\mathbf{N}$ of order M with finite alphabet A, state space $S = A^M$ and independent samples $x_{1,1}^{n_1}$ and $x_{2,1}^{n_2}$, respectively. Then if $P(a|s)$ is defined as earlier on $(X_{1,t})$ and $Q(a|s)$ is defined similarly on $(X_{2,t})$ with both probabilities positive $\forall a$ and $N_{n_i}(s) \neq 0$ for a string $s \varepsilon S$ then

(i) if $P(a|s) = Q(a|s)$, $\forall a \varepsilon A$, then $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) \to 0$ as $min(n_1, n_2) \to \infty$ and

(ii) if $\exists a \varepsilon A$ such that $P(a|s) \neq Q(a|s)$ then $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) \to \infty$ as $min(n_1, n_2) \to \infty$

Proof 5: This is shown beginning on page 871 in [2]. •

3.1 Partition Markov Model

We may now consider equivalence classes of strings in S, use these to find the Partition Markov Model, and then define a distance here. We follow **[1]** closely in the sequel.

Let (X_t) be a discrete time Markov chain, with a finite alphabet A and finite order M. Then let $S = A^M$ be the state space. Denote the string $a_m a_{m+1} ... a_n$ by a_m^n, where $a_i \varepsilon A, m \leq i \leq n$. For each $a \varepsilon A$ and $s \varepsilon S$ let $P(a|s) = Prob(X_t = a | X_{t-M}^{t-1} = s)$. We have:

Definition 6: Let (X_t) be a discrete time order M Markov chain, with a finite alphabet A and finite order M. Let $S = A^M$ be the state space. Then $s, r \varepsilon S$ are **equivalent** , denoted $s \sim_p r$, if $P(a|s) = P(a|r) \forall a \varepsilon A$.

The following is stated, without proof, in **[1]** . We provide the proof for completeness.

Proposition 6: $s \sim_p r$ is an equivalence relation.

Proof 6: **a.)**That $s \sim_p s$ is obvious; **b.)**If $s \sim_p r$ then $P(a|s) = P(a|r) \forall a \varepsilon A$, hence $P(a|r) = P(a|s) \forall a \varepsilon A$ and therefore $r \sim_p s$. Finally, **c.)**If $s \sim_p r$ and $r \sim_p q$ then $P(a|s) = P(a|r) \forall a \varepsilon A$ and $P(a|r) = P(a|q) \forall a \varepsilon A$ and therefore $P(a|s) = P(a|q) \forall a \varepsilon A$ hence $s \sim_p q$ and the proof is complete •.

We may now consider the **equivalence classes** formed by this equivalence relation. These define a **partition** $\mathcal{L}$ of the state space S.

Definition 7: The Markov chain (X_t) has partition $\mathcal{L} = \{L_1, L_2, \cdots, L_{|\mathcal{L}|}\}$ consisting of the equivalence classes defined by the equivalence relation $s \sim_p r$.

Then each class is a subset of S with the same transition probabilities. We are now able to define a distance measure between the classes.

Let x_1^n be a sample of the process (X_t) where $s \varepsilon S$, $a \varepsilon A$ and $n > M$. First, $N_n(s)$ is the number of occurrences of the string s in the sample x_1^n. That is,

$$N_n(s) = |\{t : M < t \leq n, x_{t-M}^{t-1} = s\}|$$

Then $N_n(s,a)$ is the number of occurrences of the string s followed by a in the sample x_1^n. That is,

$$N_n(s,a) = |\{t : M < t \leq n, x_{t-M}^{t-1} = s, x_t = a\}|$$

This leads to the total number of strings in L and the number of occurrences of elements into L followed by a to be, where $L\varepsilon\mathcal{L}$, respectively,

$$N_n(L) = \sum_{s\varepsilon L} N_n(s); N_n(L,a) = \sum_{s\varepsilon L} N_n(s,a)$$

Once we define what it means to have a **good** partition we can then state the distance formula in $\mathcal{L}$:

Definition 8: Let (X_t) be a discrete time order M Markov chain, with a finite alphabet A and finite order M. Let $S = A^M$ be the state space and let $\mathcal{L} = \{\mathrm{L}_1, \mathrm{L}_2, \cdots, \mathrm{L}_{|\mathcal{L}|}\}$ be a partition of S. Then

(a): $L\varepsilon\mathcal{L}$ is a **good part** of $\mathcal{L}$if $\forall s, s'\varepsilon L$,$Prob(X_t = .|X_{t-M}^{t-1} = s) = Prob(X_t = .|X_{t-M}^{t-1} = s')$ for values of $t : t > M$;

(b): $\mathcal{L}$ is a **good partition** of S if for each $i\varepsilon\{1, 2, \cdots, |\mathcal{L}|\}$, L_i satisfies (a).

Now we give the distance:

Definition 9: Let (X_t) be a discrete time order M Markov chain, with a finite alphabet A and finite order M. Let $S = A^M$ be the state space where x_1^n is a sample of the process and $\mathcal{L} = \{\mathrm{L}_1, \mathrm{L}_2, \cdots, \mathrm{L}_{|\mathcal{L}|}\}$ is a good partition of S. Letting $N_n(L_{ij}, a) = N_n(L_i, a) + N_n(L_j, a), \forall a\varepsilon A$ and $N_n(L_{ij}) = N_n(L_i) + N_n(L_j)$ then the distance in S is defined as follows:

$$d_{\mathcal{L}}(i,j) = \tfrac{1}{ln(n)} \textstyle\sum_{a\varepsilon A} \{N_n(L_i,a) ln \tfrac{N_n(L_i,a)}{N_n(L_i)} + N_n(L_j,a) ln \tfrac{N_n(L_j,a)}{N_n(L_j)} - N_n(L_{ij},a) ln \tfrac{N_n(L_{ij},a)}{N_n(L_{ij})}\}$$

This distance $d_{\mathcal{L}}(i,j)$ allows us to obtain a **metric** in the state space S as follows:

Proposition 10: Let (X_t) be a discrete time order M Markov chain, with a finite alphabet A and finite order M. Let $S = A^M$ be the state space where x_1^n is a sample of the process. Furthermore suppose that $\mathcal{L} = \{\mathrm{L}_1, \mathrm{L}_2, \cdots, \mathrm{L}_{|\mathcal{L}|}\}$ is a good partition of S for each n and for all $i, j, k\varepsilon\{1, 2, \cdots, |\mathcal{L}|\}$. Then $d_{\mathcal{L}}(i,j)$ is a **metric** in the state space S. In particular:

1.) $d_{\mathcal{L}}(i,j) \geq 0$ with equality iff $\frac{N_n(L_i,a)}{N_n(L_i)} = \frac{N_n(L_j,a)}{N_n(L_j)} \forall a\varepsilon A$;

2.) $d_{\mathcal{L}}(i,j) = d_{\mathcal{L}}(j,i)$;

3.) $d_{\mathcal{L}}(i,k) \leq d_{\mathcal{L}}(i,j) + d_{\mathcal{L}}(j,k)$.

Proof 10: The proofs of 1.) and 3.) are given in **[1]** . We include the proof of 2.) for completeness: Noting that

$N_n(L_{ij}, a) = N_n(L_i, a) + N_n(L_j, a) = N_n(L_j, a) + N_n(L_i, a) = N_n(L_{ji}, a)$ then

$$d_{\mathcal{L}}(i,j) = \tfrac{1}{ln(n)} \textstyle\sum_{a\varepsilon A} \{N_n(L_i,a) ln \tfrac{N_n(L_i,a)}{N_n(L_i)} + N_n(L_j,a) ln \tfrac{N_n(L_j,a)}{N_n(L_j)} - N_n(L_{ij},a) ln \tfrac{N_n(L_{ij},a)}{N_n(L_{ij})}\} =$$

$\frac{1}{ln(n)}\sum_{a\varepsilon A}\{N_n(L_j,a)ln\frac{N_n(L_j,a)}{N_n(L_j)}+N_n(L_i,a)ln\frac{N_n(L_i,a)}{N_n(L_i)}-N_n(L_{ji},a)ln\frac{N_n(L_{ji},a)}{N_n(L_{ji})}\}=d_{\mathcal{L}}(j,i)$ and the proof is complete •.

In Section 5 we will relate the Bayesian Information Criterion to the partition Markov model.

4 Induced Spaces

4.1 Metric Space

We follow [5] closely in this exposition:

Definition 11: A pair of objects (X,d) where X is a non-empty set and d is a function from $X \times X \to \mathbf{R}$, where $\mathbf{R}$ is the set of real numbers, is called a **metric space** provided that, $\forall x,y,z\varepsilon X$ then:

a.) $d(x,y) \geq 0$, with equality iff $x=y$;

b.) $d(x,y)=d(y,x)$;

c.) $d(x,z) \leq d(x,y)+d(y,z)$.

Examples 12: (S,d_s) and $(S,d_{\mathcal{L}})$ are metric spaces where d_s and $d_{\mathcal{L}}$ are defined as above.

Now that we have seen that (S,d_s) and $(S,d_{\mathcal{L}})$ form metric spaces we may apply all of the results that come from this fact. We have the following definitions and results, true in general for metric spaces, which will eventually allow us to consider topological spaces:

Definition 13: i) Let $(S,d_s), s\varepsilon S, x_{1,1}^{n_1}$ be as above and let $\delta>0$ be given. Then the subset of S consisting of all $x_{2,1}^{n_2}$ such that $d_s(x_{1,1}^{n_1},x_{2,1}^{n_2})<\delta$ is called the **open ball about** $x_{1,1}^{n_1}$ **of radius** δ and is denoted $B(x_{1,1}^{n_1};\delta)$.

ii) Let $(S,d_{\mathcal{L}}), L_i\varepsilon\mathcal{L}, \delta>0$ be as above. Then the subset of S consisting of all $L_j\varepsilon\mathcal{L}$ such that $d_L(i,j)<\delta$ is called the **open ball about** L_i **of radius** δ and is denoted $B(L_i;\delta)$.

Definition 14: i) Let $(S,d_s), s\varepsilon S, x_{1,1}^{n_1}$ be as above and let $\delta>0$ be given. A subset N of S is called a **neighborhood of** $x_{1,1}^{n_1}$ if there is a $\delta>0$ such that $B(x_{1,1}^{n_1};\delta)\subset N$.

ii) Let $(S,d_{\mathcal{L}}), L_i\varepsilon\mathcal{L}, \delta>0$ be as above. A subset N of S is called a **neighborhood of** L_i if there is a $\delta>0$ such that $B(L_i;\delta)\subset N$.

Lemma 15: $B(x_{1,1}^{n_1};\delta)$ and $B(L_i;\delta)$ are neighborhoods of each of their points in (S,d_s) and $(S,d_{\mathcal{L}})$, respectively.

Proof 15: This is true in general for metric spaces. See [5] for a proof. •

Definition 16: A subset O of a metric space is **open** if it is a neighborhood of each of its points.

Example 17: $B(x_{1,1}^{n_1};\delta)$ and $B(L_i;\delta)$ are open in (S,d_s) and $(S,d_{\mathcal{L}})$, respectively.

Proposition 18: A subset O of (S,d_s) or $(S,d_{\mathcal{L}})$ is an open set iff it is a union of open balls.

Proof 18: This is true in general of metric spaces. See [5] for a proof. •

Proposition 19: Let $(S,d_s), s\varepsilon S, x_{1,1}^{n_1}$ and $(S,d_{\mathcal{L}}), L_i\varepsilon\mathcal{L}$ be as above. Then, for each space respectively,

1.) The empty set is open.

2.) S is open (in the latter case as a partition of itself).

3.) If $O_1,O_2,\cdots,O_n$ are open in S then $O_1\cap O_2\cap\cdots\cap O_n$ is open.

4.) If $\{O_\alpha\}$ is any collection of open sets in S then $\cup_\alpha O_\alpha$ is open in S.

Where S represents either of the spaces.

Proof 19: This is true in general of metric spaces. See [5] for a proof. •

4.2 Topological Space

Definition 20: Let X be a non-empty set and let τ be a collection of subsets of X such that:

1.) $\phi\varepsilon\tau$,

2.) $X\varepsilon\tau$,

3.) If $O_1,O_2,\cdots,O_n\varepsilon\tau$ then $O_1\cap O_2\cap\cdots\cap O_n\varepsilon\tau$,

4.) If $\{O_\alpha\}\varepsilon\tau$ then $\cup_\alpha O_\alpha\varepsilon\tau$.

Then (X,τ) is called a **topological space** and the elements of τ are called **open sets**.

Example 21: By Proposition 19 if τ is the collection of open sets of a metric space (X,d) then (X,τ) is a topological space, induced from the metric space. These are the **metrizable** topological spaces

Example 22: Let (S,d_s) and $(S,d_{\mathcal{L}})$ be as above. Then (S,τ_s) and $(S,\tau_{\mathcal{L}})$ are topological spaces, where the context should be clear.

We may now assign to (S,τ_s) and $(S,\tau_{\mathcal{L}})$ those properties which are true of the metrizable topological spaces. For example:

Proposition 23: If a and b are distinct points of a metric space X then there are neighborhoods N_a and N_b of a and b respectively such that $N_a \cap N_b = \phi$.

Proof 23: See [5]. •

Definition 24: A subset N of X, where (X, τ) is a topological space, is called a **neighborhood** of $a \varepsilon X$ if N contains an open set containing a.

Definition 25: A topological space (X, τ) is a **Hausdorff space** if for each pair of distinct points a, b in X then there are neighborhoods N, M of a, b respectively such that $N \cap M = \phi$.

Proposition 26: All metrizable spaces are Hausdorff.

Proof 26: By Proposition 23 and the definition of metrizable spaces. •

Example 27: Let (S, τ_s) and $(S, \tau_{\mathcal{L}})$ be as above. Then (S, τ_s) and $(S, \tau_{\mathcal{L}})$ are Hausdorff spaces.

We can also introduce the **basis** for the topological space. First:

Definition 28: A collection of open sets of a metric space is called a **basis for the open sets** if each open set is the union of sets in this collection.

Example 29: The open balls in a metric space form a basis for the open sets in the metric space.

Definition 30: If (X, τ) is a topological space and $(O_\alpha)\varepsilon\tau$ then $(O_\alpha)\varepsilon\tau$ is a **basis for the open sets of** X if each open set is a union of members of (O_α).

Example 31: The open balls in (S, d_s) and $(S, d_{\mathcal{L}})$ form a basis for the open sets in (S, τ_s) and $(S, \tau_{\mathcal{L}})$, respectively.

We are now ready to apply Theorem 5 above.

Definition 32: A metric space (X, d) is **compact** if its associated topological space is compact. (For a definition of topological compactness see [5].)

Proposition 33: If (X, d) is a compact metric space then X is bounded with respect to d.

Proof 33: See [5], e.g. •

Proposition 34: The topological (metric) space (S, τ_s) is not compact if $\exists a \varepsilon A$ such that $P(a|s) \neq Q(a|s)$.

Proof 34: By Theorem 5 (ii), $d_s(x_{1,1}^{n_1}, x_{2,1}^{n_2}) \to \infty$ as $min(n_1, n_2) \to \infty$ if $\exists a \varepsilon A$ such that $P(a|s) \neq Q(a|s)$. But then there is no positive number K for which $d(,) < K$ for all entries. Hence S is not bounded with respect to d and therefore (S, d_s) is not compact, by Proposition 33. Thus (S, τ_s) is not a compact topological spaces in this scenario. •

We conclude our discussion of induced spaces with identification spaces. We again closely follow [5]. Consider the topological space $(S, \tau_{\mathcal{L}})$ defined above. Let $S/\sim_p$ be the collection of equivalence classes defined under the equivalence relation $\sim_p$ as shown in Proposition 6. Then let $\pi : S \to S/\sim_p$ which maps each $s \varepsilon S$ into its equivalence class. Then π is *onto* and we can assign to $S/\sim_p$ a topology where the open sets in $S/\sim_p$ are those sets whose inverse image under π are open $(S, \tau_{\mathcal{L}})$. Then $S/\sim_p$ is a topological **identification space** with the **identification topology**.

5 Statistical Interpretation

In their work noted above, the authors in [1] and [2] relate the Bayesian Information Criterion (BIC) of Schwarz [3] to the metrics described above. We saw that earlier in the first case and now do the same for the partition Markov model. First recall the following definitions:

$$N_n(L) = \sum_{s \varepsilon L} N_n(s); N_n(L, a) = \sum_{s \varepsilon L} N_n(s, a)$$

where $L \varepsilon \mathcal{L}$. Then $N_n(L)$ is a function of of the partition $\mathcal{L}$. So if we write $P(x_1^n)$ for $Prob(X_1^n = x_1^n)$ we get

$$P(x_1^n) = P(x_1^M) \Pi_{L \varepsilon \mathcal{L}, \mathrm{a} \varepsilon \mathrm{A}} P(a|L)^{N_n(L,a)}$$

The BIC is then defined through the *modified maximum likelihood* ML (see [3]).

$$ML(\mathcal{L}, \mathrm{x_1^n}) = \Pi_{\mathrm{L} \varepsilon \mathcal{L}, \mathrm{a} \varepsilon \mathrm{A}} (\frac{\mathrm{N_n(L, a)}}{\mathrm{N_n(L)}})^{\mathrm{N_n(L,a)}}, \mathrm{N_n(L)} \neq 0, \mathrm{L} \varepsilon \mathcal{L}$$

(Note that we had a similar formula in Section 3, which led to a definition like the one which follows.)

Definition 35: Let (X_t) be a discrete time order M Markov chain, with a finite alphabet A and finite order M. Let $S = A^M$ be the state space where x_1^n is a sample of the process. Furthermore suppose that $\mathcal{L}$ is a partition. Then

$$BIC(\mathcal{L}, \mathrm{x_1^n}) = \ln(\mathrm{ML}(\mathcal{L}, \mathrm{x_1^n})) - \frac{(|\mathrm{A}| - 1)|\mathcal{L}|}{2} \ln(\mathrm{n})$$

.

We can now relate the BIC and the metric, as a Corollary to Proposition 10. (Note that there is a similar statement as a Corollary to Theorem 4, using Definition 1.)

Corollary 36: Let (X_t) be a M order Markov chain, with a finite alphabet A. Let $S = A^M$ be the state space where x_1^n is a sample of the process. Furthermore suppose that $\mathcal{L}$ is a partition, $\{L_1, L_2, \cdots, L_{|\mathcal{L}|}\}$ such that L_i, L_j are good parts. Then, denoting $\mathcal{L}^{\mathrm{ij}}$ as the partition $\{L_1, \cdots, L_{i-1}, L_{ij}, L_{i+1}, \cdots, L_{j-1}, L_{j+1}, \cdots, L_{|\mathcal{L}|}\}$, where $L_{ij} = L_i \cup L_j$ we have

$$BIC(\mathcal{L}, \mathrm{x}_1^{\mathrm{n}}) - \mathrm{BIC}(\mathcal{L}^{\mathrm{ij}}, \mathrm{x}_1^{\mathrm{n}}) < 0 \iff \mathrm{d}_{\mathcal{L}}(\mathrm{i}, \mathrm{j}) < \frac{(|\mathrm{A}| - 1)}{2}$$

Proof 36: See Corollary 2 in [1] (or Theorem 2 in [2]). •

This Corollary provides the statistical interpretation of the distances. It was used by the authors in the proof of Theorem 4 above and later in numerous applications (see reference list).

6 Applications

In [1] the authors use their results to model internet navigation patterns. They identify strings that can be considered equivalent in terms of the next step of internet surfers, using $d_{\mathcal{L}}$ to find the minimal partition. This information can be important in determining user profiles as is desired by browser companies.

In [2], the authors from [1], plus R. Gholizadeh, use their technique to decide if 2 lines of production are equivalent, in this case the different columns used to produce fuel from sugar cane. Moreover, using the metric defined above, they are able to identify the strings that mark discrepancies between the processes. As the industry would want each column to perform similarly, this is important information for ensuring this.

We should point out that the authors mentioned above, along with others, have applied their work to many and varied areas. For example, the authors of [1], plus M. T. A. Cordeiro, J. E. Garcıa and S. L. M. Londono, have examined in [6] the Stochastic Profile of Strains of Zika utilizing the 4 bases of DNA, A = a,c,g,t, adenine (a), cytosine (c), guanine (g) and thymine (t). They find the probability of one base following another in the genomic sequence. Such an analysis may be applied to other vector-borne disease such as dengue [7], Eastern Equine Encephalitis, malaria, West Nile virus and Yellow Fever. It also may be used to study Covid-19 [8]. The author is involved in an EU Cost Action (see acknowledgements) investigating such vector-borne diseases.

This examination of base order in genomic sequences may also be applied to investigations into Z-DNA, the left-handed version of the usually right-handed molecule. It has been experimentally determined what orders of bases are likely to be in left-handed conformation. The technique in the previous paragraph can be used to find the probability of such sequences occurring. This is an area ripe for further study as it is believed that segments of Z-DNA *in vivo* can lead to deleterious effects given their different bonding properties.

Furthermore, there is a topological approach to deciding if Z-DNA is present by studying its 3-dimensional conformation, as studied by the author following [9], [10].

7 Conclusion

In this paper we extend the work of other authors in developing a Bayesian Information Criterion based consistent distance (metric) between Markovian processes. The metric space and associated topological space induced from this distance formula are presented and some properties are shown. In particular we prove that this space cannot be compact.

We then described applications of this work, some of which have already been performed, and others which follow immediately from the techniques in this paper. For example the study of base ordering in the DNA of vector-borne diseases. The paper finishes with an outline of how this work, along with topology, can be used in the study of Z-DNA.

8 Acknowledgements

This work was supported in part by an EU Cost-Action grant IMAAC (CA 16227). The study was performed by the author at The American University of Malta, who is thankful for its support.

9 References

1. Garcia JE, Gonzalez-Lopez VA. Consistent Estimation of Partition Markov Models. *Entropy* 2017:**19**: 160-175. https://doi.org/10.3390/e19040160

2. Garcia JE, Gholizadeh R, Gonzalez-Lopez VA. A BIC-based consistent metric between Markovian processes. *Appl Stochastic Models Bus Ind.* 2018:**34**:868-878. https://doi.org/10.1002/asmb.2346

3. Schwarz G. Estimating the dimension of a model. *Ann. Stat.* 1978:**6**:461-464.

4. Csiszar I, Shields PC. The consistency of the BIC Markov order estimator. *Ann. Stat.* 2000:**28**(6):1601-1619.

5. Mendelson B. Introduction to Topology. *Allyn and Bacon, Inc.* copyright 1975.

6. Cordeiro MTA, García JE, González-López VA, Londono SLM. Partition Markov model for multiple processes. *Math. Meth. Appl. Sci.* 2020:**43**:7677-7691. https://doi.org/10.1002/mma.6079

7. García JE, González-López VA. Robust Minimal Markov Model for Dengue Virus Type 3. In: Skiadas C.H., Skiadas C. (eds) Demography of Population Health, Aging and Health Expenditures. *The Springer Series on Demographic Methods and Population Analysis, Springer* . 2020: **50** .

8. García JE, González-López VA, Tasca GH. Partition Markov Model for Covid-19 Virus. *4open* 2020:**3**:11 pages. https://doi.org/10.1051/fopen/2020013

9. Bozeman JR. More Genus-One Whitehead Manifolds and Counterexamples to a Conjecture of Brown. *Proc. Jangjeon Math. Soc.* 2006:**9**:1-7.

10. Bozeman JR. Those Links in Solid Tori Which Admit a Unique Prime Decomposition. *Proc. Jangjeon Math. Soc.* 2005:**8**:1-11.