REVISTA DE MATEMÁTICA: TEORÍA Y APLICACIONES 2014 **21**(2) : 249–260
CIMPA – UCR ISSN: 1409-2433

GEODESIC DISTRIBUTION IN GRAPH THEORY: KULLBACK-LEIBLER-SYMMETRIC

DISTRIBUCIÓN GEODÉSICA EN TEORÍA DE GRAFOS: KULLBACK-LEIBLER-SIMÉTRICA

JOSÉ ALEJANDRO GONZÁLEZ* MARCOS HENRIQUE CASCONE†

Received: 4/Sep/2013; Revised: 4/Jun/2014;
Accepted: 10/Jun/2014

---

*Departamento de Estadística, Universidad de Playa Ancha. Valparaiso, Chile. E-Mail: jgonzalez@upla.cl

†IMECC, UNICAMP. Campinas, Brasil. E-Mail: marcos.cascone@gmail.com

Abstract

Kullback-Leibler information allow us to characterize a family of distributions denominated Kullback-Leibler-Symmetric, which are distance functions and, under some restrictions, generate the Jensen's equality shown by [1], in this paper denominated Jensen-Equal. On the other hand, [5] and [7] showed that graph theory gives conditions to define a new measurable space and, therefore, new distances, in particular, the distance characterized by [2], denominated Geodesic Distance. The interaction of these ideas allow us to define a new distribution, denominated Geodesic Distribution which, under graph theory as center and radius of a graph, we can to develop optimization methodologies based in probabilities of attendance. We obtain many applications and the proposal method is very adaptive. To illustrate, we apply this distribution in spatial statistics.

Keywords: Kullback-Leibler information; graph theory; geodesic distance; geodesic distribution.

Resumen

La información de Kullback-Leibler permite caracterizar una familia de distribuciones que denominamos Kullback-Liebler-Simétricas de las cuales tenemos distribuciones que son funciones de una distancia que bajo restricciones genera la igualdad en la relación de Jensen mostrados por [1], las que denominamos Jensen-Igual. Por otra parte, [5] y [7] presentan que la teoría de grafos permite definir un espacio medible y por tanto nuevas distancias, en particular la caracterizada por [2] denominada distancia Geodésica. La interacción de las dos ideas permite inducir una distribución que denominaremos Geodésica, la cual bajo técnicas de la teoría de grafos, como el centro y el radio de un grafo, permite desarrollar metodologías de optimización en función de las probabilidades de atendimiento. Obtenemos muchas áreas de aplicación y muchas adaptaciones, en las cuales, por ejemplo, aplicamos en un problema de estadística espacial.

Palabras clave: información Kullback-Leibler; teoría de grafos; distancia geodésica; distribución geodésica.

Mathematics Subject Classification: 62P99.

1 Introduction

Kullback-Leibler information, discussed in [6], measures how two informations are far from each other in the sense of likelihood. That is, if one observation came from a specific distribution, with what degree of certainty you can say that did not come from another distribution?

Let P and Q be two probability measures in a space where both measures are absolutely continuous with respect to a common measure μ. Let p and q be their densities. The Kullback-Leibler Information is given by

$$\mathcal{L}_X^{kl}(P,Q) = \int \log\left(\frac{p(x)}{q(x)}\right) p(x) d\mu x = E_p\left[\log\left(\frac{p(x)}{q(x)}\right)\right]. \quad (1)$$

In general, $\mathcal{L}_X^{kl}(P,Q)$ is not a metric since $\mathcal{L}_X^{kl}(P,Q) \neq \mathcal{L}_X^{kl}(Q,P)$. The sum $\mathcal{L}_X^{kl}(P,Q) + \mathcal{L}_X^{kl}(Q,P)$ is symmetric and it is known as Kullback-Leibler divergence. In general, divergence fails in the triangle inequality, so it is not a metric ([9]).

In this paper, we study conditions for which the distribuitions are Kullback-Leibler-Symmetric and with these results, we will utilize the geodesic distance to define a distribution in graphs.

2 Preliminary

2.1 Conditions for the Kullback-Leibler-Symmetric

Definition 1 *A function $f(\cdot)$ is called Jensen-Equal if $E[f(x)] = f[E(x)]$.*

Let $m(x;\theta)$ be a metric such that $m(x;\theta) = f(x-\theta)$, and let X be a random variable with mean given by θ. Then the function $g(x;\theta) = m(x;\theta_1) - m(x;\theta)$, $\theta, \theta_1 \in \mathbb{R}$, is a Jensen-Equal function. For instance, we have two cases, as follows.

Case 2 *Consider $m(x;\theta) = (x-\theta)^2$, that is, the quadratic metric.*

Case 3 *Consider $m(x;\theta) = |x-\theta|$, that is, the euclidean metric.*

Proof of Case 2: Let X be a random variable with mean θ and $g(x;\theta) = (x-\theta_1)^2 - (x-\theta)^2$. Then

$$\begin{aligned} E\left[(x-\theta_1)^2 - (x-\theta)^2\right] &= E\left(x^2 - 2x\theta_1 + \theta_1^2 - x^2 + 2x\theta - \theta^2\right) \\ &= E\left(-2x\theta_1 + \theta_1^2 + 2x\theta - \theta^2\right) \\ &= E\left(x\left(2\theta - 2\theta_1\right) + \theta_1^2 - \theta^2\right). \end{aligned}$$

Since $x\,(2\theta - 2\theta_1) + \theta_1^2 - \theta^2$ is a linear function, then we can show that

$$E\left[(x-\theta_1)^2 - (x-\theta)^2\right] = (\theta_1 - \theta)^2 .$$

On the other hand,

$$[E\,(x) - \theta_1]^2 - [E\,(x) - \theta]^2 = (\theta_1 - \theta_1)^2 - (\theta_1 - \theta)^2 = (\theta_1 - \theta)^2 .$$

So $E\left[(x-\theta_1)^2 - (x-\theta)^2\right] = [E\,(x) - \theta_1]^2 - [E\,(x) - \theta]^2$, and $g(x;\theta)$ is a Jensen-Equal function. ■

Proof of Case 3: Let X be a random variable with mean θ and $g(x;\theta) = |x - \theta_1| - |x-\theta|$. Then we have that

$$\begin{aligned}
E\,(|x-\theta_1| - |x-\theta|) &= \begin{cases} E\,(x) - \theta_1 - E\,(x) + \theta & \text{, if } \theta \leq x \text{ and } \theta_1 \leq x \\ -E\,(x) + \theta_1 - E\,(x) + \theta & \text{, if } \theta \leq x \text{ and } \theta_1 \geq x \\ -E\,(x) + \theta_1 + E\,(x) - \theta & \text{, if } \theta \geq x \text{ and } \theta_1 \geq x \\ (E\,(x) - \theta_1) + E\,(x) - \theta & \text{, if } \theta \geq x \text{ and } \theta_1 \leq x \end{cases} \\
&= \begin{cases} -\theta 1 + \theta & \text{, if } \theta \leq x \text{ and } \theta_1 \leq x \\ \theta_1 + \theta & \text{, if } \theta \leq x \text{ and } \theta_1 \geq x \\ \theta_1 - \theta & \text{, if } \theta \geq x \text{ and } \theta_1 \geq x \\ -\theta_1 - \theta & \text{, if } \theta \geq x \text{ and } \theta_1 \leq x \end{cases} \\
&= |\theta_1 - \theta|. \qquad (2)
\end{aligned}$$

On the other hand,

$$|E\,(x) - \theta_1| - |E\,(x) - \theta| = (|\theta - \theta_1| - |\theta - \theta|) = |\theta - \theta_1| . \qquad (3)$$

By (2) and (3), $g(x,\theta)$ is a Jensen-Equal function. ■

Generally, $\mathcal{L}_X^{kl}(P,Q)$ is not a metric since $\mathcal{L}_X^{kl}(P,Q) \neq \mathcal{L}_X^{kl}(Q,P)$. We will now characterize the functions that define a metric.

Theorem 4 *If $g(m(x;\theta)) = k \cdot \exp\{w \cdot m(x;\theta)\} = f(x;\theta)$, where $f(x;\theta)$ is a density function such that $E(X) = c$, $c \in \mathbb{R}$ and $(k,w) \in \mathbb{R} \times \mathbb{R}^-$, then $\mathcal{L}_X^{kl}(P,Q)$ is a metric.*

Proof. To show that $\mathcal{L}_X^{kl}(P,Q)$ is a metric, we need to verify three conditions:

1. $\mathcal{L}_X^{kl}(P,Q) \geq 0$;
2. $\mathcal{L}_X^{kl}(P,Q) = \mathcal{L}_X^{kl}(Q,P)$, and
3. $\mathcal{L}_X^{kl}(P,R) \leq \mathcal{L}_X^{kl}(P,Q) + \mathcal{L}_X^{kl}(Q,R)$.

To show the condition 1, note that

$$\begin{aligned}
E\left(\log\left(\frac{f(x,\theta_1)}{f(x,\theta_2)}\right)\right) &= E\left(\log\left(\frac{g(m(x,\theta_1))}{g(m(x,\theta_2))}\right)\right) \\
&= E\left(\log\left(\frac{k\cdot\exp(w\cdot m(x,\theta_1))}{k\cdot\exp(w\cdot m(x,\theta_2))}\right)\right) \\
&= wE\{m(x,\theta_1)-m(x,\theta_2)\},
\end{aligned}$$

where $\theta_1, \theta_2 \in \mathbb{R}$ and the function $m(x,\theta_1)-m(x,\theta_2)$ is Jensen-Equal. Hence

$$\begin{aligned}
wE\{m(x,\theta_1)-m(x,\theta_2)\} &= w\{m[E(x),\theta_1]-m[E(x),\theta_2]\} \\
&= w[m(\theta_1,\theta_1)-m(\theta_1,\theta_2)] \\
&= -w[m(\theta_1,\theta_2)] \geq 0. \qquad (4)
\end{aligned}$$

Similarly, $E\left(\log\left(\frac{f(x,\theta_2)}{f(x,\theta_1)}\right)\right) = -w[m(\theta_2,\theta_1)]$ and as $m(\cdot,\cdot)$ is a metric, we have that $-w[m(\theta_2,\theta_1)] = -w[m(\theta_1,\theta_2)]$.

Thus, $\mathcal{L}_X^{kl}(P,Q) = \mathcal{L}_X^{kl}(Q,P)$. With the same argument about $m(\cdot,\cdot)$, we have that $m(\theta_1,\theta_3) \leq m(\theta_1,\theta_2)+m(\theta_2,\theta_3)$ and, therefore, $-w[m(\theta_1,\theta_3)] \leq -w[m(\theta_1,\theta_2)] - w[m(\theta_2,\theta_3)]$.

Since $E\left(\log\left(\frac{f(x,\theta_i)}{f(x,\theta_j)}\right)\right) = -w[m(\theta_i,\theta_j)]$,

$$E\left(\log\left(\frac{f(x,\theta_1)}{f(x,\theta_3)}\right)\right) \leq E\left(\log\left(\frac{f(x,\theta_1)}{f(x,\theta_2)}\right)\right) + E\left(\log\left(\frac{f(x,\theta_2)}{f(x,\theta_3)}\right)\right) \qquad (5)$$

and (5) shows that $\mathcal{L}_X^{kl}(P,R) \leq \mathcal{L}_X^{kl}(P,Q) + \mathcal{L}_X^{kl}(Q,R)$.

Hence, as the Conditions 1, 2 and 3 are satisfied, $\mathcal{L}_X^{kl}(P,Q)$ is a metric. ■

Theorem 4 allow us to define a family of distributions denominated Kullback-Leibler-Symmetric. For instance, the normal and double exponential (Laplace) distributions belongs to Kullback-Leibler-Symmetric family.

2.2 Graph theory and geodesic distance

In this section, we will define some characteristics of graphs that will be used throughout the paper.

Definition 5 *A graph G is a pair $(V(G), A(G))$, where $V(G)$ is a non-empty finite set of objects called vertices and $A(G)$ is a set (possibly empty) of unordered pairs of distinct vertices of G called edges or lines ([4]).*

The edge $a = \{v, w\}$ is said to *join* the vertices v and w. If $a = \{v, w\}$ is an edge of a graph, then v and w are called *adjacent vertices*, while v and a are incident, as are w and a. Furthermore, if a_1 and a_2 are distinct edges of G incident with a common vertex, then a_1 and a_2 are *adjacent edges*. From now, we will denote an edge by vw or wv. Also, the degree of a vertex v in a graph G is the number of edges of G incident with v.

The *cardinality* of the vertex set of a graph G is called the order of G and is denoted by $O(G)$, or more simply o, and the cardinality of its edge is the size of G, often denoted by $T(G)$ or simply t. Therefore, a graph has order o and size t.

It is customary to define or describe a graph by means of a diagram in which each vertex is represented by a point and each edge is represented by a line segment or curve joining the points corresponding to v and w.

A graph G, with a vertices set $V(G) = \{v_1, v_2, \ldots, v_p\}$ can also be described by means of matrices. Then, $A(G) = [a_{ij}]$, where

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in A(G) \\ 0 & \text{if } v_i v_j \notin A(G). \end{cases}$$

Such matrix is called adjacency matrix of G.

Another matrix is called *incidence matrix*, that also represent a graph. Then $B(G) = [b_{ij}]$ where

$$b_{ij} = \begin{cases} 1 & \text{if } v_i \text{ and } a_j \text{ are incidents} \\ 0 & \text{otherwise.} \end{cases} .$$

One important remark of a graph is when it is complete. A graph is said to be complete if for a number n of vertices, the maximum number of edges defined is $n(n-1)/2$, that characterizes a complete graph, denoted by K_n ([3] ;[8]). The next definition is extremely important to de development of this paper.

Definition 6 (Geodesic Distance) *In the graph theory, the distance between two vertices v and w in a graph G is the number of edges of the shortest path that connects the vertices v and w, called Geodesic Distance and denoted by $dg(v, w)$.*

Let $G(A)$ and $G(V)$ be the set of edges and vertices, respectively, of a graph G, vw the edge generate by v and w and $G(A) = G_1(A) \cup G_2(A)$ such that $G_1(A) \cap G_2(A) = \emptyset$. Thus, if $v \in G_1(V)$ and $w \in G_2(V)$, then $dg(v, w) = \infty$. In other words, if there is no path connecting two vertices (that is, if the vertices belongs to different sub-graphs unconnected), then the distance between these two graphs is defined infinite. The eccentricity ϵ of a vertex v is the biggest geodesic distance between v and any other vertex. It can be thought of as how much a node is far from the most distant node in the graph.

The property of eccentricity allow us to define the *radius* and the *diameter* of a graph. The radius of a graph is the minimum eccentricity among the vertices of the graph. The diameter of a graph is the maximum eccentricity of any vertex of the graph. That is, the diameter is the greatest distance between any pair of vertices. To find a diameter of a graph, first find the shortest path among each pair of vertices. The greatest distance of anyone path is the diameter of the graph.

Definition 7 *A central subgraph of G is the graph that has n vertices with degree α and the smallest eccentricity. It will be denoted by $G_{Cn\alpha}$.*

Definition 8 *The attractor graph of G is the graph that has the n vertices with greatest degree, denoted by $G_{A_{n\beta}}$, where β represent the greatest degree of the graph G.*

3 Proposal

The Kullback-Leibler-symmetric provide us tools to analyse the regularly behavior of the densities which depends of a measure Jensen-Equal. Therefore, the idea of planning dependent distributions based on distances follows in a natural way, and given the wide field of application, the geodesic distance was the choice to represent the measure of this proposal.

Definition 9 *If G is a graph with order c, then the Central Geodesic density function is given by*

$$f(v, \alpha, k) = k \exp\{-dg(v,_n \theta_\alpha\},$$

where ${}_n\theta_\alpha$ is $G_{Cn\alpha}$, $dg(v, \theta_\alpha)$ is the Geodesic Distance between the vertex v and the graph ${}_n\theta_\alpha$ and k is the normalizing constant, $k \in [1, \infty)$.

Definition 10 *If G is a graph with order c, then the Attractor Geodesic density function is given by*

$$f(v, \beta, k) = k \exp\{-dg(v,_n \theta_\beta\},$$

where ${}_n\theta_\beta$ is G_{A_β}, $dg(v, \theta_\beta)$ is the Geodesic Distance between the vertex v and the graph ${}_n\theta_\beta$ and k is the normalizing constant, $k \in [1, \infty)$.

If $m(h(x), \theta)$ is a metric such that $h(x) = z$ is a linear function of x and X is a random variable with mean θ, then the function $g(x;\theta) = m(z, \theta_1) - m(z, \theta)$ is a Jensen-Equal function, $\theta, \theta_1 \in \mathbb{R}$. The lognormal model is a function of the metric $(\log(x)-\theta)^2$. However $\log(x)$ is not linear function of x and, therefore, it is not Jensen-Equal. Hence, metrics of the form $m(h(x), \theta)$, such that $h(x) = z$ is linear function of x, induce a distribution function and if m is a continuous or discrete measure, m induce a quantity or density function given by

$$f(x, \theta) = k \exp\{w[m(z, \theta)]\},$$

where $w \in \mathbb{R}^-$, $k \in \mathbb{R}^+$ and $\theta \in \mathbb{R}$.

Some properties of Geodesic distribution defined above are given as follows:

1. $dg(v_{i,n}\, \theta_{\alpha,\beta}) \in \mathbb{N}_0$.
2. $\sum_{i=0}^{\infty} \exp\{-dg(v_{i,n}\, \theta_{\alpha,\beta})\} = 1/k$.
3. $\sum_{i=0}^{\infty} \exp\{-dg(v_{i,n}\, \theta_{\alpha,\beta})\} = \sum_{i=0}^{\alpha-1} c_i \exp\{-i\} = c_1 + \sum_{i=1}^{\alpha-1} c_i \exp\{-i\}$.
4. $\sum_{i=0}^{\alpha-1} c_i = c$.
5. $c_1 = e$, where e is the order of ${}_n\theta_{\alpha,\beta}$.

It is important to note that working with central graphs tends to equate the probability of meeting the vertices.

4 Application

The Kirke Channel is located on XII Region of Magallanes and Chilean Antarctic. It is continuation of the Morla Vicuña channel and forms part of the maritime access to Puerto Natales, capital city of Ultima Esperanza province. This channel was navigated by kawesqar people since 6000 years ago until XX century, because inhabited its coast. Also, this channel separates south coast of Diego Portales island in the north side of Vicuña Mackenna peninsula. Diego Portales island is very charming for its vegetation and two lines of hills that form it. Marks the boundary between the Pampa region and island region. Among its hills, the most important are Diego and Portales, ending in a sharp and snow-capped peaks, both about one thousand two hundred meters high.

Its coast are clean except in three points marked by kelp bass. The main difficulty to the navigation are the strong current and its 50 meters wide of navigable sea in some parts. The following is the map of the vicinity of the Kirke channel.

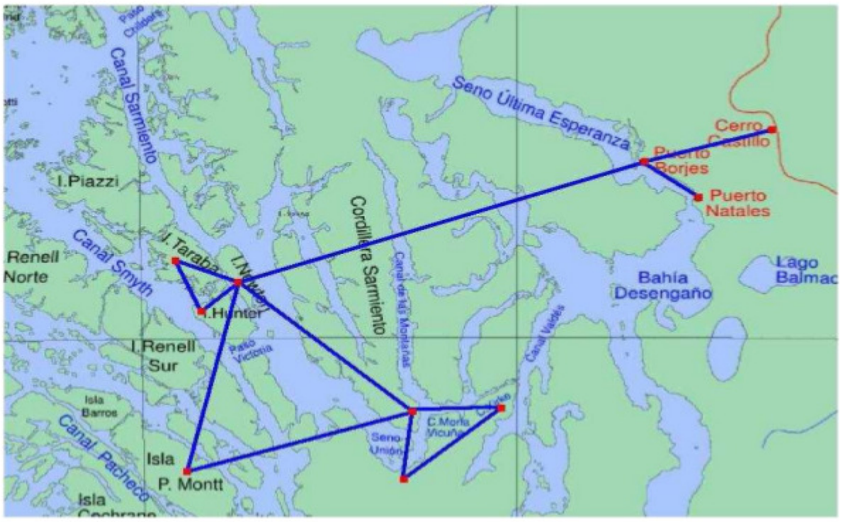

Figure 1: Vicinity of Kirke channel.

Problem

The goal is to build a hospital that meets the needs of the ten villages. We will assume that all these villages has the same population and the edges as the same weight. The population of the villages think that the better place to build the hospital is on Morla Vicuña channel for best location and needs a formal authorization for this opinion.

The question is: Do you agree with these people? Why? If not, which village do you believe is better to build the hospital? Why?

The abstract solution to this problem is given by the Figure 2, where C represents the Morla Vicaña channel.

Here, we will consider C the graph to be treated by the villages interest, and after we will apply the attractor graph methodology to compare the results and answer the question.

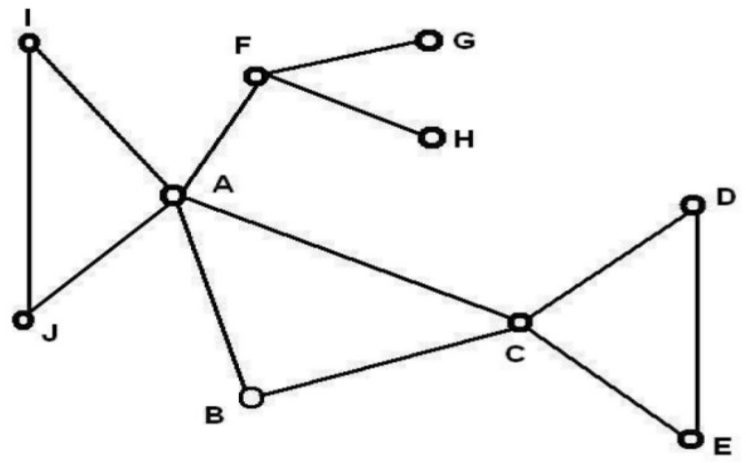

Figure 2: Vicinity of Kirke channel.

The eccentricities are

$$A \to 4,\ B \to 3,\ C \to 4,\ D \to 5,\ E \to 5,$$
$$F \to 5,\ G \to 6,\ H \to 6,\ I \to 5,\ J \to 5,$$

and therefore the eccentricity of G is $Ecc(G) = 6$.

The vertex C has eccentricity 4, so we say that $X_C \sim Geodesic({}_1\theta_4)$. The attractor graph is constituted by the vertex A, where the degree of this vertex is $\beta = 5$. As in the vertex C, we will denote X_A by $X_A \sim Geodesic({}_1\theta_5)$.

In the first case, the distribution function is given by

Table 1: First case CDF.

x	**G**	**F**	**A**	**B**	**C**
$P(X=x)$	0,016	0,042	0,115	0,115	0,312
x	**D**	**E**	**I**	**J**	**H**
$P(X=x)$	0,115	0,115	0,115	0,042	0,016

In the second case, the distribution function is given by

Table 2: Second case CDF.

x	**D**	**E**	**B**	**C**	**A**
$P(X=x)$	0,040	0,040	0,109	0,109	0,296
x	**F**	**I**	**J**	**G**	**H**
$P(X=x)$	0,109	0,109	0,109	0,040	0,040

5 Conclusions

Thus, by the methodology here developed, Vicuña Morla channel is not the best place to build a hospital because the probabilities of meeting the villages near Vicuña Morla has higher variability than the probabilities obtained by the attractor graph method.

This methodology tends to homogenize the probability of meeting, so the best place to build a hospital is Newton Island (vertex A).

In future works, the constant w might assume different weights according to the situation, for instance, car traffic, blood pressure and others.

Acknowledgements

The authors are grateful to Coordination of Improvement of Higher Education Personnel, CAPES, for the financial support. The first author is grateful to Experimental Laboratory of Mathematical Knowledge, LAB[E]SAM and Laboratory Data Analysis and Statistical Knowledge, LADYSE. University of Playa Ancha and “Concurso regular de proyectos de investigación año 2013. Cambios sustantivos en la formación de profesores de la Universidad de Playa Ancha de Ciencias de la Educación. Procesos formativos de excelencia para impactar en

el rendimiento escolar y fortalecimiento social de las comunidades educativas, UPA 2013".

References

[1] Beckenbach, E. (1937) "Generalized convex functions", *Bulletin of American Mathematical Society*, **43**(6): 363–371.

[2] Bouttier, J.; Di Francesco, P; Guitter, E. (2003) "Geodesic distance in planar graphs", *Nuclear Physics B* **663**(3): 535–567.

[3] Chartrand, G.; Oellermann, O.R. (1993) *Applied and Algorithmic Graph Theory*. McGraw-Hill, New York.

[4] Chartrand, G.; Lesniak, L. (1996) *Graphs and Digraphs*. Chapman & Hall/CRC, New York.

[5] Chung, F.; Horn, P.; Lu, L. (2012) "Diameter of random spanning trees in a given graph", *Journal of Graph Theory* **69**(3): 223–240.

[6] Lachos, V.H.; Bandyopadhyay, D.; Dey, D.K. (2011) "Linear and non-linear mixed-effects models for censored HIV viral loads using normal/independent distributions", *Biometrics* **67**(4): 1594–1604.

[7] Montenegro, E.; Cabrera, E. (2001) "Attractors points in the autosubstitution", *Proyecciones (Antofagasta)* **20**(2): 193–204.

[8] Montenegro, E.; González, J.; Cabrera, E.; Nettle A.; Robres, R. (2010) "Graphs r-polar spherical realization", *Proyecciones (Antofagasta)* **29**(1): 31–39.

[9] Schervish, M.J. (1995) *Theory of Statistics*. Springer, New York.