Progress in Probability
79

Daniel Hernández-Hernández
Florencia Leonardi
Ramsés H. Mena
Juan Carlos Pardo Millán
Editors

Advances in Probability and Mathematical Statistics

CLAPEM 2019, Mérida, Mexico

CIMAT

Birkhäuser

Birkhäuser

Progress in Probability

Volume 79

Progress in Probability is designed for the publication of workshops, seminars and conference proceedings on all aspects of probability theory and stochastic processes, as well as their connections with and applications to other areas such as mathematical statistics and statistical physics.

All manuscripts are peer-reviewed to meet the highest standards of scientific literature. Interested authors may submit proposals by email to the series editors or to the relevant Birkhäuser editor listed under "Contacts."

More information about this series at http://www.springer.com/series/4839

Daniel Hernández-Hernández • Florencia Leonardi •
Ramsés H. Mena • Juan Carlos Pardo Millán
Editors

Advances in Probability and Mathematical Statistics

CLAPEM 2019, Mérida, Mexico

Birkhäuser

Editors
Daniel Hernández-Hernández
Department of Probability and Statistics
Centro de Investigación en Matemáticas
Guanajuato, Mexico

Florencia Leonardi
Instituto de Matemática e Estatística
Universidade de São Paulo
São Paulo, Brazil

Ramsés H. Mena
Departamento de Probabilidad
IIMAS-UNAM
Mexico City, Mexico

Juan Carlos Pardo Millán
Department of Probability and Statistics
Centro de Investigación en Matemáticas
Guanajuato, Mexico

ISSN 1050-6977 ISSN 2297-0428 (electronic)
Progress in Probability
ISBN 978-3-030-85324-2 ISBN 978-3-030-85325-9 (eBook)
https://doi.org/10.1007/978-3-030-85325-9

Mathematics Subject Classification: 93E20, 90C40, 60J65, 60G51, 60F15, 60F10; Primary: 62C10, 60G57, 62F15, 99Z99; Secondary: 00A00, 62B10

This book is published under the imprint Birkhäuser, www.birkhauser-science.com, by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The present volume contains contributions of the XV Latin American Congress of Probability and Mathematical Statistics (CLAPEM, by its acronym in Spanish), held at Merida, Mexico during December 2–6, 2019.

Endorsed by the Bernoulli Society, this event is the official meeting of the *Sociedad Latinoamericana de Probabilidad y Estadística Matemática* (SLAPEM) and it is the major event in probability and statistics in the region. It gathers an important number of researchers and students, predominantly from Latin America, serving as an ideal forum to discuss and to disseminate recent advances in the field, as well as to reveal the future of our profession.

Over nearly 40 years, the CLAPEMs have greatly contributed to the development of probability and statistics by promoting collaborations in the region as well as with the rest of the world. Previous editions were held in Caracas (1980, 1985, 2009), Montevideo (1988), Ciudad de México (1990), San Pablo (1993), Viña del Mar (1995, 2012), Córdoba (1998), La Habana (2001), Punta del Este (2004), Lima (2007), Cartagena (2014), and San José (2016).

On this occasion, the congress gathered scholars from over 20 countries and included a wide set of topics on probability and statistics. The scientific program included four plenary talks delivered by Gerard Ben Arous, Sourav Chatterjee, Thomas Mountdford, and Judith Rousseau. The event also benefited from eight semi-plenary talks given by Pablo Ferrari, Michele Guindani, Chris Holmes, Jean Michel Marin, Lea Popovic, and Fernando Quintana. The program also included two courses: " Hierarchical Bayesian Modeling and Analysis for Spatial BIG Data " by Sudipto Banerjee and " Sharpness of the phase transition in percolation " by Vincent Tassion, 10 thematic sessions, 21 contributed sessions, and several contributed talks and poster presentations.

The volume begins with the chapter by Andrade, Calvillo, Manrique, and Treviño where the authors present a probabilistic analysis of random interval graphs associated with randomly generated instances of the data delivery on a line problem (or DDLP). Angel and Spinka consider the infinite random geometric graph on a circle of circumference L, which is a random graph whose vertex set is given by a dense countable set in such circle, and find a dependency behavior on the rationality of L. The asymptotic behavior of four binary classification methods, when the dimension of the data increases and the sample sizes of the classes are fixed, are studied by Bolivar-Cime. A class of transport distances based on the Wassertein distances for random vectors of measures is considered by Catalano, Lijoi, and Prünster. The latter leads to a new measure of dependence for completely random vectors, and the quantification of the impact of hyperparameters in notable models for exchangeable time-to-event data. Gil-Leyva studies the construction of random discrete distributions, taking values in the infinite dimensional simplex, by means of latent random subsets of the natural numbers, which are then applied to construct Bayesian non-parametric priors. The connection between generalized entropies based on a certain family of α-divergences and the class of some predictive distributions is studied by Gutiérrez-Peña and Mendoza. A class of discrete-time stochastic controlled systems composed by a large population of N interacting individuals is considered by Higuera-Chan. The problem is studied by means of the so-called mean field model. Kouarfate, Kouritzin, and Mackay provide an explicit weak solution for the $3/2$ stochastic volatility model which is used to develop a simulation algorithm for option pricing purposes. Finally, León and Rouault revisit Wschebor's theorems on the a.s. convergence of small increments for processes with scaling and stationarity properties and then apply such results to deduce that large deviation principles are satisfied by occupation measures.

In summary, the high quality and variety of these chapters illustrate the rich academic program at the XV CLAPEM. It is worth noting that all papers were subject to a strict refereeing process with high international standards. We are very grateful to the referees, many leading experts in their own fields, for their careful and useful reports. Their comments were addressed by the authors, allowing to improve the material in this volume.

We would also like to extend our gratitude to all the authors whose original contributions appear published here as well as to all the speakers of the XV CLAPEM for their stimulating talks and support. Their valuable contributions encourage the interest and activity in the area of probability and statistics in Latin America.

We hold in high regard the editors of the series Progress in Probability: Davar Khoshnevisan, Andreas E. Kyprianou, and Sidney I. Resnick for giving us the opportunity to publish the symposium volume in this prestigious series.

Special thanks go to the Universidad Autónoma de Yucatán and its staff for its great hospitality and for providing excellent conference facilities. We are indebted to Rosy Dávalos, whose outstanding organizational work permitted us to concentrate mainly in the academic aspects of the conference.

The XV CLAPEM as well as the publication of this volume would not have been possible without the generous support of our sponsors and the organizing institutions: Bernoulli Society; Centro de Investigación en Matemáticas; Consejo Nacional de Ciencia y Tecnología; Facultad de Ciencias–UNAM; Gobierno de Yucatán; Google; Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas–UNAM; Instituto de Matemáticas–UNAM; Universidad Autónoma de Chapingo; Universi-

dad Autónoma Metropolitana; Universidad Autónoma de Yucatán; and Universidad Juárez Autónoma de Tabasco.

Finally, we hope the reader of this volume enjoys learning about the various topics treated as much as we did editing it.

Guanajuato, Mexico	Daniel Hernández-Hernández
São Paulo, Brazil	Florencia Leonardi
Mexico City, Mexico	Ramsés H. Mena
Guanajuato, Mexico	Juan Carlos Pardo Millán

Contents

Asymptotic Connectedness of Random Interval Graphs in a One Dimensional Data Delivery Problem

Caleb Erubiel Andrade Sernas, Gilberto Calvillo Vives, Paulo Cesar Manrique Mirón, and Erick Treviño Aguilar

Abstract In this work we present a probabilistic analysis of random interval graphs associated with randomly generated instances of the Data Delivery on a Line Problem (DDLP) (Chalopin et al., Data delivery by energy-constrained mobile agents on a line. In Automata, languages, and programming, pp. 423–434. Springer, Berlin, 2014). Random Interval Graphs have been previously studied by Scheinermann (Discrete Math 82:287–302, 1990). However, his model and ours provide different ways to generate the graphs. Our model is defined by how the agents in the DDLP may move, thus its importance goes beyond the intrinsic interest of random graphs and has to do with the complexity of a combinatorial optimization problem which has been proven to be NP-complete (Chalopin et al., Data delivery by energy-constrained mobile agents on a line. In Automata, languages, and programming, pp. 423–434. Springer, Berlin, 2014). We study the relationship between solvability of a random instance of the DDLP with respect to its associated interval graph connectedness. This relationship is important because through probabilistic analysis we prove that despite the NP-completeness of DDLP, there are classes of instances that can be solved polynomially.

Keywords Connectedness analysis · Data delivery problem · Mobile agents · Random interval graph

1 Introduction

The research presented in this work is in the intersection of several disciplines, Probability Theory, Computer Science, Operations Research, and Graph Theory. So, in this introduction we define the problem, and provide the basic concepts of computational complexity and random graphs that are needed to make the work

C. E. Andrade Sernas · G. Calvillo Vives · P. C. Manrique Mirón · E. Treviño Aguilar (✉)
Unidad Cuernavaca del Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico City, Mexico
e-mail: erick.trevino@im.unam.mx

D. Hernández-Hernández et al. (eds.), *Advances in Probability and Mathematical Statistics*, Progress in Probability 79, https://doi.org/10.1007/978-3-030-85325-9_1

self-contained. We also include several references for those that may want to go deeper in the study of these subjects.

1.1 The Data Delivery on a Line Problem

The production of inexpensive, simple-built, mobile robots has become a reality nowadays, to the point that a swarm of mobile agents can be used for different tasks. A practical application would be, for instance, to use a swarm of drones to explore a cave and produce a map of it by collecting and sharing geospatial data whenever two drones meet. Or the use of a network of drones to deliver packages to customers from retail stores or courier services. The Data Delivery Problem is a mathematical abstraction of such scenarios and has been studied a lot lately, see e.g., [1, 3, 4, 8, 9, 12]. Here we deal with a specific Data Delivery Linear Problem (DDLP), namely where agents (robots) are constrained to move in a line. In this version, a set of n energy-constrained mobile agents are placed on positions $0 \leq x_i \leq 1, i = 1, \ldots, n$ on the unit interval and have ranges $\rho_i > 0$ (denoting the maximum length of a walk for each agent), the question is whether there is an order in which the mobile agents should be deployed to pick up the data at a point s called the source, and collectively move it to a predetermined destination called the target $t > s$. The first agent, in the order found, moves to the source, picks up the data and move it to the right according to its capacity. The second agent moves to the point where the first agent is, takes the data and move further to the right where a third agent comes to take over and so on until the data arrives, if possible, at its destination t. Observe that there are two cases for the movement of an agent. If the agent is to the left of the position d where the data is, it moves always to the right. First to pick up the data and then to move it further to the right. If it is to the right of d, then it has to move first to the left to reach d and then to the right as far as it can. In both cases, the agent covers the range $[x_i - a\rho_i, x_i + (1 - 2a)\rho_i]$ where $a = 0$ for the first case and $a = (x_i - d)/\rho_i$ for the second. This observation is key to define graphs associated to the problem. Observe also that the DDLP is a decision problem; that is to say, the answer is yes or no. In this framework now we can talk about the computational complexity of the problem.

This problem was introduced by Chalopin et al. [9], and it was shown to be NP-complete, although for instances where all input values are integers they gave a quasi-pseudo-polynomial time algorithm. In Sect. 2 we show how this problem is equivalent to a graph theoretical problem which in turn is analyzed using random graphs in Sect. 3. So let us first briefly recall some computational complexity and graph theoretical concepts.

1.2 Computational Complexity

This field has several roots: The seminal ideas of Jack Edmonds [13] about good algorithms and good characterizations; the foundational work of Stephen Cook [11] and Dick Karp [17], to name a few. The field is now represented by one of the millennium problems: *Is P=NP?*. A systematic treatment of the subject can be found in [15], [2].

A decision problem is a collection of instances (propositions) each of which is true or false. The DDLP is one such decision problems. Each instance is of the form: The set of agents $Q = \{(x_i, \rho_i), i = 1, \ldots, n\}$ can move the data from s to t. This proposition is true or false. The problem is to decide for each instance which is the correct answer. A decision problem is said to belong to the *class P* if there is an algorithm (Turing Machine) that decides correctly which is the answer and runs in polynomial time, which means that the number of steps the corresponding Turing Machine has to perform is bounded by a polynomial in the size of the instance (usually measured in bits). For the DDLP, the size of an instance is the number of bits required to store Q, s and t. A famous Decision Problem in the class P is to decide if a given set of linear inequalities has a solution. A decision problem belongs to the *class NP* if for every instance with answer YES, there is a polynomial algorithm to verify that the answer is correct. The DDLP is NP since for an affirmative answer it suffices to show the sequence in which the agents are used and check that effectively they move the data from s to t. This, of course can be done very efficiently. A decision problem S belongs to the class *NP-complete* if it is NP and any other NP problem can be reduced polynomially to S. Cook provided the first NP-complete problem (satisfiability) and then Karp [17] added a bunch of combinatorial problems to that class. Chalopin showed that DDLP is a NP-complete problem. While question *is P=NP?* is open, we do not know if the NP complete problems can be solved by polynomial time algorithms. At present, the generalized belief is that $P \neq NP$. If this is so, NP problems will remain as difficult problems, including the DDLP. One of the drawbacks of this theory is that it does not recognize clearly that a large amount of an NP-complete problem can be solved efficiently. The recognition of that fact has motivated the use of probability to asses fringes of “easy” instances and thus isolate the really hard instances of a problem. A good example of this methodology is in [6]. Our work follows that path. In Sect. 5 we interpret the results presented in Sect. 3 in this sense.

1.3 Graphs

For the purpose of this work it is sufficient to deal with simple graphs and so we omit the adjective. A graph consists of a set V whose elements are called vertices and a collection of subsets of V of cardinality 2 which we denote by E and call them edges. If a vertex v belongs to an edge e we say that v is an extreme of e. Clearly every edge has two extremes. If u and v are the two extremes of an edge we say that they are adjacent. A simple path (or path) is a sequence $v_0, e_1, v_1, e_2, v_2, \ldots, v_{n+1}, e_n, v_n$ in which all vertices v_i, $i = 1, \ldots, n$ are different and v_i and v_{i+1} are the extremes of edge e_i. Such path is said to connect vertices v_0 and v_n; when $v_0 = v_k$, the graph is a cycle, denoted by C_k. A graph is connected if for every pair of vertices u and v of the graph, there exists a path connecting them. Given a set V of real closed intervals in the real line we can construct a graph in the following way: The set of vertices is V and two vertices are adjacent (form an edge) if they intersect as intervals. A graph constructed in this way is called an **interval graph**. The graphs that can be constructed in this way are a small part of all possible graphs. For example, any cycle C_k with $k > 3$ is not an interval graph. However they are nice graphs to work with because they have remarkable properties. There are several ways in which a collection of intervals can be defined. We are interested in two of them.

Definition 1 The symmetric model of a interval graph is generated by intervals I_i defined by its center x_i and its radius ρ_i in the form $I_i = [x_i - \rho_i, x_i + \rho_i]$. The interval graph associated to the symmetric model will be denoted by $G_S(Q)$ where $Q = \{(x_i, \rho_i), i = 1, \ldots n\}$.

The **asymmetric model** is obtained from instances of DDLP considering $I_i = [x_i - a_i \rho_i, x_i + (1 - 2a_i)\rho_i]$ defined in Sect. 1.1. It will be denoted by $G_A(Q, a)$, where again $Q = \{(x_i, \rho_i), i = 1, \ldots n\}$ and a is a vector in the unit box of $\mathbb{R}^n$.

A standard reference in graph theory is [7].

1.4 Random Interval Graphs

Random graphs were firstly introduced by Gilbert [16], but the work of Erdős and Rényi [14] set the basis for the study of their evolution (a comprehensive study on random graphs can be found in [6]). Later on, Cohen studied the asymptotic probability that a random graph is a unit interval graph [10], and Scheinermann replicated the ideas of Erdős and Rényi to study the evolution of random interval graphs [18].

To continue our discussion, we introduce a probabilistic framework, so let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $\mathcal{F}$ a σ-algebra on the set of scenarios Ω and $\mathbb{P}$ a probability measure on $\mathcal{F}$. All random variables and our asymmetric model will be defined there.

Scheinerman [18] obtained random interval graphs, using the symmetric model, by defining two random variables for each interval, the centers $\{x_i\}_{i=1}^n$ and the radii of the intervals $\{\rho_i\}_{i=1}^n$, the centers with uniform distribution in [0,1] and the radii also uniform in the interval $[0, r]$ with $r < 1$, so intervals are constructed as $[x_i - \rho_i, x_i + \rho_i]$. We use the asymmetric model fixing the parameters $\{a_i\}_{i=1}^n$, $a_i \in [0, 1]$ and considering $2n$ independent random variables, the locations $\{x_i\}_{i=1}^n$ (all identically distributed uniformly in $[0, 1]$), and the ranges $\{\rho_i\}_{i=1}^n$ (all identically distributed uniformly in $[0, r]$). The intervals are $[x_i - a_i \rho_i, x_i + (1 - 2a_i)\, \rho_i]$, which are substantially different to the intervals of the symmetric model. First they are asymmetric, second they are shorter as specified by the parameter a_i which represents the percentage of a mobile agent's energy used to move backwards to pick up the load before going forward with it, thus it determines the behavior of the agent. The asymptotic analysis presented in this work is analogous to the one of Scheinermann's, but here we consider an asymmetric model which is more general. Moreover, it is related to a NP-complete combinatorial problem which makes the new model interesting from the point of view of computational complexity, a fact that we show in this work.

2 Graph Theoretical Formulation of the Data Delivery on a Line Problem

In this section we show how the DDLP is equivalent to an existence problem for interval graphs. The transformation from one another is completely general and does not assume that the data have been randomly generated. The equivalent graph theoretical decision problem is the following: Given a finite collection $Q = \{(x_i, \rho_i), i = 1, \ldots n\}$ of points in $\mathbb{R}^2$ decide if there exists a vector $a = (a_1, \ldots, a_n)$ such that the interval graph $G_A(Q, a)$ defined by the asymmetric model is connected. This problem will be called the *Existential Connectedness Problem* (ECP).

We say that two decision problems are equivalent if for every instance of one there is an instance of the other such that both have an affirmative answer or both have a negative answer.

Theorem 1 *DDLP and ECP are equivalent.*

Proof First we will show that for every instance (Q, s, t) of DDLP there exists an instance of ECP whose solution conforms with the one of the DDLP instance. First, discard all points of Q such that either $x_i + \rho_i < s$ or $x_i - \rho_i > t$ then add two new points $(s, 0)$ and $(t, 0)$. Call this new set of points Q' which defines an instance of the ECP. A positive answer to this instance comes with a vector a^* such that $G_A(Q', a^*)$ is a connected graph. So there exists a path T in $G_A(Q', a^*)$ from the vertex defined by $(s, 0)$ to the vertex defined by $(t, 0)$. The interior vertices of T correspond to points of Q and so to agents in the DDLP instance. The order in which the vertices of T are traversed from s to t define the order in which the agents have to be deployed. The adjacency of consecutive vertices of T guarantee that the corresponding agents can get in touch to pass the data from one to the next. So the existence of an affirmative solution of ECP translates into a solution of DDLP.

It remains to be proven that if the solution to the ECP instance is negative so is the answer to the DDLP. The contrapositive of this is that any YES answer to an instance of DDLP translates into an affirmative answer of ECP. To prove this let

(Q, s, t) be an instance of DDLP with a true answer and a sequence $r_{i_1}, \ldots, r_{i_k}$ of agents that carry the data from s to t. This behaviour of the agents in the sequence define a set of parameters $a_{i_1}, \ldots, a_{i_k}$. For the rest of the agents, which do not play any role in the transportation of the data, let $a_i = 0$ if $x_i < t$ and $a_i = 1$ if $x_i \geq t$. We claim that the graph $G_A(Q, a)$ is connected. The parameters $a_{i_1}, \ldots, a_{i_k}$ define a set of intervals, a path T' in $G_A(Q, a)$, that cover the segment $[s, t]$. We will show that every other interval intersects T'. For an agent r_i such that $x_i < t$ the corresponding interval is $[x_i, x_i + \rho_i]$, if $x_i \geq s$, then x_i is in $[s, t]$ and so it intersects some interval of T'; if $x_i < s$, then since irrelevant agents have been removed $x_i + \rho_i \geq s$ and therefore the interval intersects T' too. The case $x_i \geq$ t is resolved similarly. □

The next result is a direct consequence to Theorem 1. We omit the details of the proof.

Corollary 1 *ECP is NP-complete*

Theorem 1 shows that DDLP, which is NP-complete, is reducible to ECP. Moreover the reduction is polynomial since to transform a DDLP given by (Q, s, t) to a ECP given by Q' it is only needed to compute which robots (x_i, ρ_i) satisfy $x_i - \rho_i < s$ or $x_i + \rho_i > t$; and to include the points $(s, 0)$, $(t, 0)$. This can be done in linear time assuming arithmetic operations are performed in constant time. The equivalence between DDLP and ECP allow us to deal with the graph theoretical formulation and use the results and ideas of Random Graphs in order to obtain some asymptotic results in the ECP and therefore the DDLP problems. Specifically, parameters' domain can be partitioned into regions, one in which the problems can efficiently be solved. The relevance of the connectedness of an interval graph in relation to the solvability of its associated DDLP instance cannot be overstated. That is, if for any DDLP instance (Q, s, t) there exists a vector a in the unit box such that $G_A\left(Q', a\right)$ is connected, then the DDLP instance is solvable, the converse is certainly not true, as shown in the counterexamples of Fig. 1. On the other hand, given a DDLP instance (Q, s, t), if for all possible vectors a in the unit box none of the associated interval graphs $G_A\left(Q', a\right)$ is connected, it is then assured that the instance is not solvable. This last statement is summarized in the following corollary.

Corollary 2 *Given a DDLP instance (Q, s, t) if its associated symmetric interval graph $G_S\left(Q'\right)$ is disconnected, the instance is not solvable.*

Proof For every vector a in the unit box we have that

$$[x_i - a y_i, x_i + (1 - 2a)\, y_i] \subset [x_i - y_i, x_i + y_i]\,.$$

Now, recall that $Q' = Q \cup \{(s, 0)\,, (t, 0)\}$, so if $G_S\left(Q'\right)$ is disconnected, it means that for every vector a in the unit box $G_A\left(Q', a\right)$ is disconnected as well. This means that ECP is not solvable, so by Theorem 1 DDLP is not solvable. □

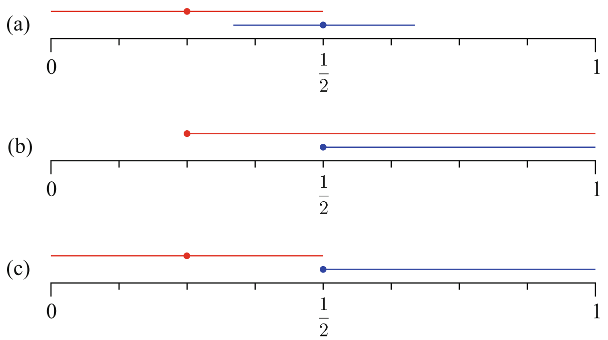

Fig. 1 For this example we take $Q = \left\{\left(\frac{1}{4}, \frac{3}{4}\right), \left(\frac{1}{2}, \frac{1}{2}\right)\right\}$. In panel (**a**) we have $a = \left(\frac{1}{3}, \frac{1}{3}\right)$, in (**b**) $a = (0, 0)$, and in (**c**) $a = \left(\frac{1}{3}, 0\right)$. In the first two cases $G_A\left(Q', a\right)$ is disconnected, but in the third case it is connected and therefore solvable

3 Probabilistic Analysis of $G_A(Q, a)$

We denote as usual the degree of a vertex v_i by $d(i)$. We write $v_i \not\sim v_j$ when there is no edge joining the vertices v_i and v_j. We denote by $v_i^+ = x_i + (1-2a)\rho_i$ (resp. $v_i^- = x_i - a\rho_i$) the right (resp. left) boundary of v_i. In this section we analyze the connectedness of random interval graphs of the form $G_A(Q, a)$. The approach that we follow is to fix a and analyze the asymptotic behavior of the connected components of $G_A(Q, a)$. In order to do so, we first introduce random variables that count the connected components of a random graph $G_A(Q, a)$. In this section and the sequel we will specialize to the case in which a is a scalar, that is $a_i = a$ for all i. Furthermore, we assume $a \in [0, 1/2]$. Estimations with this specification will already yield lower bounds for the probability of solvability due to Theorem 1 which is our main goal.

For each i, $i = 1, \ldots, n$ let

$$X_i := \begin{cases} 1 & \text{if the right end point of } v_i \text{ is contained in no other interval } v_j,\ j \neq i, \\ 0 & \text{otherwise.} \end{cases}$$

The family of random variables $X_1, \ldots, X_n$ indicates where a connected component ends. For example, assume that $X_1 = X_2 = 1$ and $X_i = 0$ for $i > 2$. Assume without loss of generality that the interval v_1 is on the left of the right end point of v_2. Since $X_1 = 1$ the right end point of v_1 is not included in any other interval, implying that all the intervals on the left of v_1^+ are disjoint from all other intervals on its right. Thus, there are at least two components. In order to see that there are exactly two components, assume that there is one more. For this component denote by v_j the interval such that v_j^+ attains the supremum over all right ends of intervals in the components. Then, necessarily $X_j = 1$ with $j > 2$. This is a contradiction. Then, there exists exactly two components.

We also define

$$X(n) := \sum_{i=1}^{n} X_i$$

is the random variable that counts the number of connected components of $G_A(Q, a)$. Note that $\mathbb{P}(X(n) = 0) = 0$. The distribution of $X(n)$ depends on the parameter r. However, we do not write this explicitly in order not to overload notation. In our first asymptotic result, we let the numbers of vertices n goes to infinite while at the same time shrinking the range of the intervals (the parameter r). The theorem gives the right trade-off for those processes in order to have connectedness with high probability.

Theorem 2 *Let* $\beta := \frac{2}{1-a}$ *and* $r(n) := \beta \frac{1}{n}(\log(n) + c)$. *Then,*

$$\liminf_{n\to\infty} \mathbb{P}(X(n) = 1) \geq 1 - e^{-c}. \tag{1}$$

Proof Denote by $v_i^+ = x_i + (1 - 2a)\rho_i$ the right point of the interval v_i. Let

$$B(i) := \{\omega \mid v_i^+(\omega) \in \cup_{k\neq i} v_k\}$$

be the event in which the right boundary of the interval v_i is included in at least one of the other intervals. The event $\{X(n) = 1\}$ satisfies

$$\{X(n) = 1\} = \bigcup_{k=1}^{n} \bigcap_{i\neq k} B(i).$$

Hence

$$\mathbb{P}(X(n) = 1) = \mathbb{P}\left(\bigcup_{k=1}^{n} \bigcap_{\substack{i=1\\ i\neq k}}^{n} B(i)\right) \geq \mathbb{P}\left(\bigcap_{i=2}^{n} B(i)\right).$$

Moreover

$$1 - \mathbb{P}\left(\bigcap_{j=2}^{n} B(j)\right) = \mathbb{P}\left(\left\{\bigcap_{j=2}^{n} B(j)\right\}^c\right) \leq \sum_{j=2}^{n} \mathbb{P}\left(B^c(j)\right) = (n-1)\mathbb{P}\left(B^c(2)\right).$$

Hence:

$$\mathbb{P}(X(n) = 1) \geq 1 - (n-1)\mathbb{P}\left(B^c(2)\right). \tag{2}$$

So now we estimate $\mathbb{P}(B^c(2))$. The right end point v_i^+ is a random variable and we denote by μ it probability distribution function under P. Note that the distribution does not depend on i. Then,

$$\mathbb{P}\left(B^c(2)\right) = \int_0^{1+(1-2a)r} (1 - \mathbb{P}(t \in v_2))^{n-1} \mu(dt),$$

due to the independency of the intervals. For $a < \frac{1}{2}$ the random variable v_i^+ is the sum of two independent uniform distributions and then, we can easily see that μ is a distribution concentrated on $[0, 1 + r(1-2a)]$ with density

$$f(x) = \begin{cases} \frac{x}{r(1-2a)} & [0, r(1-2a)] \\ 1 & [r(1-2a), 1] \\ \frac{-x+1+r(1-2a)}{r(1-2a)} & [1, 1+r(1-2a)] \\ 0 & otherwise. \end{cases}$$

For $t \in [0, 1 + r(1-2a)]$ let $h(t, x, \rho)$ be the indicator function of $\{x - a\rho \leq t\} \cap \{x + (1-2a)\rho \geq t\}$ and $g(t, \rho) := \left\{(t + a\rho) \wedge 1 - (t - (1-2a)\rho)^+ \wedge 1\right\}$. We have

$$\begin{aligned} \mathbb{P}(t \in v_2) &= \frac{1}{r} \int_0^r \int_0^1 h(t, x, \rho)\mu(dx)d\rho \\ &= \frac{1}{r} \int_0^r g(t, \rho)d\rho. \end{aligned}$$

For $t \in [0, (1-2a)r]$ the function g simplifies to $g(t, \rho) = t + a\rho - (t - (1-2a)\rho)1_{\{t>(1-2a)\rho\}}$. Hence

$$\begin{aligned} \mathbb{P}(t \in v_2) &= t + \frac{1}{2}ar - \frac{1}{r}\left[t\rho - \frac{1}{2}(1-2a)\rho^2\right]_0^{\frac{t}{1-2a}} \\ &= t + \frac{1}{2}ar - \frac{1}{2r(1-2a)}t^2. \end{aligned}$$