Searching for a Trail of Evidence in a Maze

Ery Arias-Castro[1]*, Emmanuel J. Candès[2]†, Hannes Helgason[2] and Ofer Zeitouni[3] ‡

1) Department of Mathematics, University of California, San Diego, CA 92093

2) Applied and Computational Mathematics, Caltech, Pasadena, CA 91125

3) Department of Mathematics, University of Minnesota, Minneapolis, MN 55455, and Technion, Haifa 32000, Israel

January 24, 2007

Abstract

Consider a graph with a set of vertices and oriented edges connecting pairs of vertices. Each vertex is associated with a random variable and these are assumed to be independent. In this setting, suppose we wish to solve the following hypothesis testing problem: under the null, the random variables have common distribution $N(0,1)$ while under the alternative, there is an unknown path along which random variables have distribution $N(\mu,1)$, $\mu > 0$, and distribution $N(0,1)$ away from it. For which values of the mean shift μ can one reliably detect and for which values is this impossible?

This paper develops detection thresholds for two types of common graphs which exhibit a different behavior. The first is the usual regular lattice with vertices of the form

$$\{(i,j) : 0 \le i,\ -i \le j \le i \text{ and } j \text{ has the parity of } i\}$$

and oriented edges $(i,j) \to (i+1, j+s)$ where $s = \pm 1$. We show that for paths of length m starting at the origin, the hypotheses become distinguishable (in a minimax sense) if $\mu_m \gg \sqrt{\log m}$, while they are not if $\mu_m \ll \log m$. We derive equivalent results in a Bayesian setting where one assumes that all paths are equally likely; there the asymptotic threshold is $\mu_m \approx m^{-1/4}$. We obtain corresponding results for trees (where the threshold is of order 1 and independent of the size of the tree), for distributions other than the Gaussian, and for other graphs. The concept of predictability profile, first introduced by Benjamini, Pemantle and Peres, plays a crucial role in our analysis.

Keywords. Detecting a chain of nodes in a network, minimax detection, Bayesian detection, predictability profile of a stochastic process, martingales, exponential families of random variables.

*Partially supported by National Science Foundation grant DMS-0603890

†Partially supported by National Science Foundation grants ITR ACI-0204932 and CCF515362. E. C. would like to thank the Centre Interfacultaire Bernoulli of the Ecole Polytechnique Fédérale de Lausanne for hospitality during June and July 2006

‡Partially supported by National Science Foundation grant DMS-0503775

1 Introduction

This paper discusses the model problem of detecting whether or not in a given network, there is a chain of connected nodes which exhibit an "unusual behavior." Suppose we are given a graph G with vertex set V and a random variable X_v attached to each node $v \in V$. In that sense, this is a graph indexed process. We observe a realization of this process and would like to tell whether all the variables at the nodes have the same behavior in the sense that they are all sampled from a common distribution F_0 or whether there is a path in the network, i.e. a chain of consecutive nodes connected by edges, along which the variables at the nodes have a different distribution F_1. In other words, can one tell whether hidden in the background noise, there is a chain of nodes that stand out?

Suppose for example that F_0 is the standard normal distribution whereas F_1 is a normal distribution with mean 0.1 and variance 1. In a situation where the number of nodes along the path we wish to detect is comparably small, the largest values of X_v are typically off this path. Can we reliably detect the existence of such a path? More generally, how subtle an effect can we detect? In this paper, we attempt to provide quantitative answers to such questions by investigating asymptotic detection thresholds; values of the mean shift at which detection is possible and values at which detection by any method whatsoever is impossible.

Detection thresholds depend, of course, on the type of graphs under consideration and we propose the study of two representative graphs which are, in some sense, far from each other as well as emblematic: regular lattices and trees. We introduce them next.

- **Regular lattice in dimension 2.** Our first graph is a regular lattice with nodes

  $$V = \{(i, j) : 0 \leq i, \, -i \leq j \leq i \text{ and } j \text{ has the parity of } i\},$$

  and with oriented edges $(i, j) \to (i + 1, j + s)$ where $s = \pm 1$. We call $(0, 0)$ the origin of the graph. A path in the graph is represented in Figure 1.

- **Complete binary tree.** Our second model is the oriented regular binary tree. The nodes in the tree are of the form

  $$V = \{(i, j) : 0 \leq i, \, 0 \leq j < 2^i\}.$$

  and with oriented edges $(i, j) \to (i + 1, 2j + s)$ where $s \in \{0, 1\}$. We call $(0, 0)$ the origin of the graph. A path in the tree is represented in Figure 2.

Note that even though the number of paths of length m in both graphs are the same, the number of nodes are widely different; about $m^2/2$ for the lattice, and 2^m for the binary tree.

We call $\mathcal{P}$ the set of paths in the graph starting at the origin and $\mathcal{P}_m$ those of length m. (In this paper, we define the length of a path as the number of vertices the path visits.) We attach a random variable X_v to each node v in the graph. We observe $\{X_v : v \in V\}$ and consider the following hypothesis testing problem:

- Under H_0, all the X_v's are i.i.d. $N(0, 1)$.

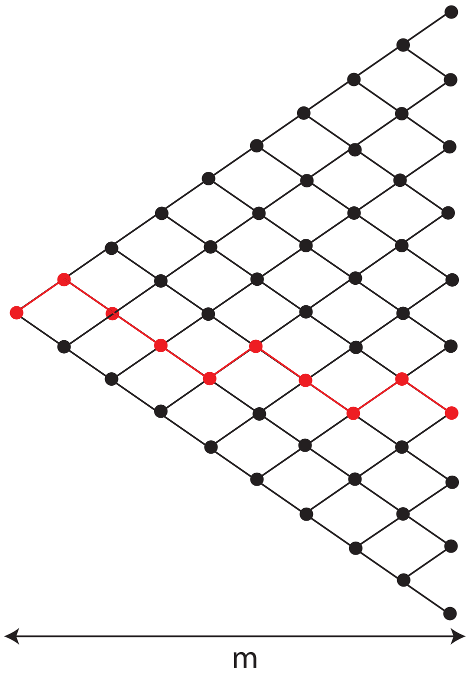

Figure 1: Representation of a path (in red) in the graph.

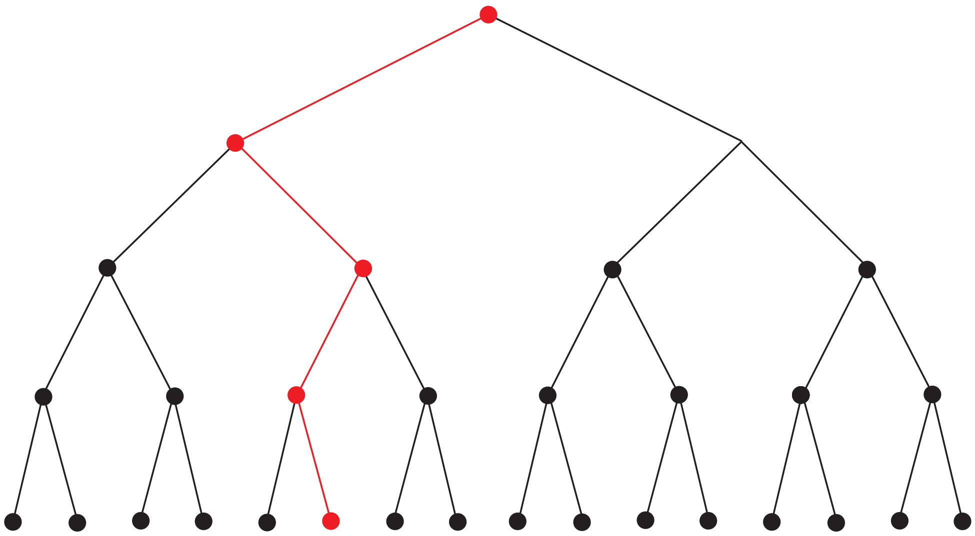

Figure 2: Representation of a path (in red) in the tree.

- Under $H_{1,m}$, all the X_v's are independent; there is an unknown path $p \in \mathcal{P}_m$ along which the X_v's are i.i.d. $N(\mu_m, 1)$, $\mu_m > 0$, while they are i.i.d. $N(0, 1)$ away from the path.

In plain English, we would like to know whether there is a path along which the mean is elevated.

1.1 Motivation

While this paper is mainly concerned with the study of fundamental detection limits, our problem is in fact motivated by applications in various fields and especially in the area of signal detection.

Suppose we are given very noisy data of the form

$$y_i = S_i + z_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where (S_i) are sampled values of a signal of interest and (z_i) is a noise term. Based on the observations (y_i), one would like to decide whether or not a signal is hiding in the noise. That is, we would like to test whether $S = 0$ or not. Suppose further that the signal is completely unknown, and does not depend on a small number of parameters. In image processing, the signal S might be the indicator function of a general shape or of a curve we wish to detect [3]. In signal processing, the signal may be a chirp, a high frequency wave with unknown and rapidly changing oscillatory patterns [11].

In these situations, we cannot hope to generate a family of candidate signals that would provide large correlations with the unknown signal as the number of such candidates would be exponentially large in the signal size. In response, recent papers [11, 14] have proposed a very different approach in which the family of candidate signals actually corresponds to a path in a network. We briefly explain the main idea. In most situations, it is certainly possible to generate a family of templates $(\phi_v)_{v \in V}$ which provide good *local* correlations with the signal of interest, e.g. over shorter time intervals. Any signal of interest could then be closely approximated by a chain of such templates. Here, a chain is a path in a graph G with nodes $v \in V$ indexed by our templates and rules for connecting templates with the following property: any consecutive sequence of templates in the graph must correspond to a meaningful signal; that is, a signal one might expect to observe (for instance, imagine connecting linear segments to approximate smooth curves). Now calculate a Z-score for each template and call it X_v. For simplicity, assume that $X_v \sim N(\mu_1, 1)$ if the templates matches the signal S locally and $X_v \sim N(\mu_0, 1)$ otherwise. Assume $\mu_1 > \mu_0$. Then the signal detection problem is this: is there a path along which the mean of the Z-scores is slightly elevated?

To make things a little more concrete, suppose the unknown signal $S(t)$ is a chirp of the general form $A(t) \exp(i\lambda\varphi(t))$, where $A(t)$ is a smooth amplitude, $\varphi(t)$ is a smooth phase function and λ is a large base frequency. Roughly speaking, a chirp is an oscillatory signal with "instantaneous frequency" given by the derivative of the phase, i.e. $\lambda\varphi'(t)$. Here, one might use as templates chirplets of the form $\phi_v(t) \propto 1_{I_v}(t) \exp(i(a_v t^2/2 + b_v t))$ which are supported on the time interval I_v and assume the linear instantaneous frequency $a_v t + b_v$. Such templates provide a local quadratic approximation to the unknown phase function $\lambda\varphi(t)$ (or a local linear approximation to the unknown instantaneous frequency) and can exhibit high correlations with the unknown signal provided that the discretization of the chirplet parameters is sufficiently fine. The chirplet graph [11] then connects pairs of chirplets supported on contiguous time intervals by imposing a certain kind of continuity of the instantaneous frequency in such a way that a path represents a chirping signal with a piecewise linear instantaneous frequency which obeys a prescribed regularity criterion. Given the data vector y (1.1), one would then compute all the chirplet coefficients $X_v = \langle y, \phi_v \rangle$ of y, and testing whether there is signal or not amounts to testing whether all the node variables X_v in the chirplet graph have mean 0 or whether there is a path along which the mean is nonzero.

Although the signal detection problem motivates the theoretical study presented in this paper, the problem of detecting a path in a network seems to represent a fundamental abstraction as many modern statistical detection problems can reasonably be formulated in this way. We give a few examples with the mere goal of stimulating the reader's imagination:

1. *Monitoring the environment and sensor networks.* In [26], water quality in a network of streams is assessed by performing a chemical analysis at various locations along the streams. As a result, some locations are then marked as problematic. We may view the set of all tested locations as nodes and connect pairs of adjacent nodes located on the same stream, thereby creating a tree (although not a regular tree). We then assign to each node the value 1 or 0 according to whether the location is problematic or not. A possible model would assume that the variables are Bernoulli which take on the value 1 with probability equal to p_0 when the location is normal and p_1 when it is anomalous. One can then imagine that one would like to detect a path (or a family of paths) downstream of a certain location, see [26].

2. *Finding structures in the distribution of galaxies.* It is of interest to find structures, and in particular filaments, in the distribution of galaxies [25, 28] based on observational data representing the positions of galaxies in the sky as enumerated in the Las Campanas redshift catalogue and the Sloan Digital Sky Survey. A possible approach would bin the galaxies in a 3D array with the 3 dimensional lattice as a model of a graph and a statistical model might assume that the galaxy counts within each cell are Poisson distributed with intensity λ. Under the uniform background hypothesis, the intensity would be λ_0 whereas a filament model would assert the existence of chains of connected cells with an intensity $\lambda_1 > \lambda_0$. In this context, one would be interested in testing whether the data provides evidence against the uniform background hypothesis. For related ideas, see [3, 8].

3. *Detection of attacks in a network.* Suppose we have a computer network which is subject to attacks [22]. For instance, a certain computer on the network may be infected by a virus which can then spread to other machines. Elements affected by the virus may exhibit an abnormal behavior (e.g. a loss of performance, violations of specific rules, etc.) which can be hard to detect on a single computer. However, we do not know which network route might be infected, and the specific sites of attacks might change randomly with time so that we never get a very definite indication that we are seeing a loss of performance in any part of the network. Here, we might still want to detect the presence of a virus; in other words, of a small fraction of infected nodes with a slightly abnormal behavior.

4. *Gene detection.* Suppose we have a series of microarray experiments and wish to tell whether or not groups of genes are differently expressed. For instance, we could have a very large set of genes measured on a few microarrays under two different experimental conditions, such as control and treatment. In this setup, we could imagine having available a two-sample t-statistic X_v for each gene. One could declare significant those genes whose t-statistic exceeds

a given threshold but this may not be a very powerful strategy [15]. In fact, one could test for the significance of a group genes, where each candidate group is a chain of connected nodes in a gene network [30]. Genes in the network could be connected if they are in the same biological pathway in the cell, or if they have close DNA sequences, or are close together on the cell's chromosome. The idea here is that the gene expression levels of those genes of interest may be buried in the noise, but that one could perhaps detect their presence by borrowing strength across the network.

We emphasize that in the examples above, one might be interested in detecting anomalous paths and/or more generally, anomalous sets of connected vertices.

1.2 Peek at the results

The optimal detection threshold discussed above is the minimum value of $\mu = \mu_m$ which allows us to reliably tell whether or not there is a path which does not follow the null distribution. This value depends on the criterion used for judging the quality of the decision rule, and statistical decision theory essentially offers two paradigms: the Bayesian and the minimax approach. We study them both.

Consider the minimax paradigm first. Recall that a test T_m is a $\{0,1\}$-valued, measurable function of the collection $\{X_v\}_{v \in V}$. The risk of a test T_m is defined as

$$\gamma(T_m) = \mathbf{P}(\text{Type I}) + \sup_{p \in \mathcal{P}_m} \mathbf{P}_p(\text{Type II}). \tag{1.2}$$

Throughout, we write $\mathbf{P}_0$ for the law of $\{X_v\}$ under H_0, and $\mathbf{P}_{1,p}$ for the law of the same variables under $H_{1,m}$ with path $p \in \mathcal{P}_m$. With these notations, Type I and II are shorthands for errors of Type I and II. In longhand,

$$\mathbf{P}(\text{Type I}) = \mathbf{P}_0(T_m = 1)\,, \quad \mathbf{P}_p(\text{Type II}) = \mathbf{P}_{1,p}(T_m = 0)\,.$$

We will say that a sequence of tests (T_m) is *asymptotically powerful* if

$$\lim_{m \to \infty} \gamma(T_m) = 0\,,$$

and *asymptotically powerless* if

$$\liminf_{m \to \infty} \gamma(T_m) \geq 1\,.$$

When there exists an asymptotically powerful sequence of tests we say that reliable detection is possible; when all sequences of tests are asymptotically powerless we say that detection is (essentially) impossible.

1.2.1 The regular lattice

We first consider the regular lattice in dimension 2.

Theorem 1.1. *Consider the regular lattice in dimension 2. Suppose that* $\mu_m (\log m)^{1/2} \to \infty$ *as* $m \to \infty$*. Then there is a sequence of tests which is asymptotically powerful. On the other hand, suppose that* $\mu_m \log m (\log \log m)^{1/2} \to 0$ *as* $m \to 0$*. Then every sequence of tests* (T_m) *is asymptotically powerless.*

Theorem 1.1 states that one can detect a path as long as $\mu_m \gg (\log m)^{-1/2}$ while this is impossible if $\mu_m < (\log m)^{-(1+\epsilon)}$ for each $\epsilon > 0$ provided that m is sufficiently large. The reader will note the discrepancy between the lower and the upper bound, which we will comment upon in the concluding section.

It turns out that the detection level is radically different in a Bayesian framework where one assumes that all paths are equally likely. For a prior π on $\mathcal{P}_m$, namely on paths of length m, the corresponding risk of a test T_m is now defined as

$$\gamma_\pi(T_m) = \mathbf{P}(\text{Type I}) + \mathbf{E}_\pi \mathbf{P}_p(\text{Type II}), \tag{1.3}$$

where $\mathbf{E}_\pi$ stands for the expectation over the prior path distribution, namely, when the path p is drawn according to π. We adapt the same terminology as before and say that (T_m) is asymptotically powerful if $\gamma_\pi(T_m) \to 0$ and powerless if $\liminf \gamma_\pi(T_m) \geq 1$. The Bayes test associated with π is of course optimal here. The theorem below shows that, under the uniform prior on paths, the optimal Bayesian detectability threshold is about $m^{-1/4}$.

Theorem 1.2. *Consider the regular lattice in dimension 2 and assume the uniform prior on paths. If* $\mu_m m^{1/4} \to \infty$ *as* $m \to \infty$*, then the Bayes risk tends to 1. Conversely, if* $\mu_m m^{1/4} \to 0$ *as* $m \to 0$*, the Bayes risk tends to 0.*

Roughly speaking, if the anomalous path is chosen uniformly at random, one can asymptotically detect it as long as the intensity along the path exceeds $m^{-1/4}$ while no method whatsoever can detect below this level.

Both results indicate that it is possible to detect an anomalous path event when $\mu_m \to 0$ (sufficiently slowly). Note that while one can certainly reliably detect in such circumstances, it may be impossible to tell which sequence of nodes the anomalous path is traversing. This is an example of a situation where detection is possible but estimation may not be.

1.2.2 The binary tree

We are now interested in the complete binary tree.

Theorem 1.3. *If* $\mu_m = \mu \geq \sqrt{2 \log 2}$*, there is a sequence of tests that is asymptotically powerful. On the other hand, if* $\mu_m = \mu < \sqrt{2 \log 2}$*, there is no sequence of tests that is asymptotically powerful. Moreover, if* $\mu_m \to 0$ *as* $m \to \infty$*, then every sequence of tests is asymptotically powerless.*

Notice that there is no sharp threshold phenomenon here, in the sense that the minimax risk does not converge to 1 if $\mu_m = \mu < \sqrt{2 \log 2}$. For example, the risk of the test which rejects the null hypothesis for large values of the variable at the root node is bounded away from 1 for any $\mu > 0$.

For any graph and under the normal mean model, consider the generalized likelihood ratio test (GLRT) which is the test rejecting the null for large values of $M_m := \max\{X_p : p \in \mathcal{P}_m\}$, where X_p is the sum of the node variables along the path p:

$$X_p = \sum_{v \in p} X_v. \tag{1.4}$$

Then the proof of Theorem 1.3 shows that the GLRT achieves the minimax threshold in that it has asymptotically full power when $\mu > \sqrt{2 \log 2}$. In this sense, the GLRT rivals the Bayes test under the uniform prior on paths, which by symmetry is minimax.

1.3 Innovations and related work

In the regular graph model, the number of variables needed to describe the path is m while the total number of nodes or observations is about $m^2/2$. Hence, the topic of this paper fits in the broad framework of nonparametric detection as the object we wish to detect is simply too complex to be reduced to a small number of parameters. Because the theory and practice of detection has been centered around parametric models in which the generalized likelihood ratio test has played a crucial role (see the literature on scan statistics, matched filters and deformable templates to name a few equivalent terms used in various fields of science and engineering [2, 18, 23, 29]), methods and results for nonparametric detection are comparably underdeveloped. Against this background, we will first provide some evidence showing that the generalized likelihood ratio test does not perform very well in our nonparametric setup. Our work also differs from the important literature on nonparametric detection in that it does not assume that the unknown object we wish to detect lies in a traditional smoothness class such as Sobolev or Besov classes or belongs to an ℓ_p-ball or some related geometric body, see the book by Ingster and Suslina [21] and the multiple references therein. In fact, our model, techniques and results have nothing to do with this literature and hence, our paper contributes to developing the important area of nonparametric detection in what appears to be a new direction. In fact, we are not familiar with statistical theory posing a problem as a graph detection problem and giving precise detection quantitative bounds; it has come to our attention, however, that Berger and Peres have very recently considered problems which are mathematically closely related to our framework but with a different motivation.

Our paper also bears some connections with the theory and practice of multiple hypothesis testing. Indeed, we are interested in situations where testing at each node separately offers little or no power so that we need to combine information from different nodes. Because the anomalous nodes are located on a path, the search naturally involves testing over paths. There are many such paths, however, and in this sense our problem resembles that of testing many hypotheses (one hypothesis test would be whether the mean along a specified path is zero or not).

1.4 Organization of the paper

The paper is organized as follows. In Section 2, we study the detection problems over the triangular array and prove our results about the minimax and Bayesian detection thresholds, namely, Theorems 1.1 and 1.2. In Section 3, we prove the detection thresholds for the binary tree. In

Section 4, we extend our results to exponential distributions at the nodes and in Section 5 to other distributions and other graphs. In Section 6, we report on numerical simulations which complement our theoretical study. Finally, we conclude with Section 7 where we comment on our findings and discuss open problems.

2 The Regular Lattice

Throughout, for positive sequences (a_m), (b_m), we write $a_m \asymp b_m$ if the ratio a_m/b_m is bounded away from zero and infinity. Also, we occasionally drop subscripts wherever there is no ambiguity to lighten the notation.

2.1 Bayesian detection

We assume the uniform distribution over all paths, denoted π. Equivalently, the distribution of the unknown path is that of an oriented symmetric random walk. We write $\mathbf{P}_\pi(\cdot) = \mathbf{E}_\pi \mathbf{P}_{1,p}(\cdot)$. As is well-known, the test minimizing the risk (1.3) is the Neyman-Pearson test which rejects the null if and only if the likelihood ratio $L_m(X) = d\mathbf{P}_\pi(X)/d\mathbf{P}_0(X)$ exceeds 1 (the subscript m refers here to the size of the problem). Here, the likelihood ratio is given by

$$L_m(X) = 2^{-(m-1)} \sum_{p \in \mathcal{P}_m} e^{\mu X_p - m\mu^2/2}, \tag{2.1}$$

where X_p is defined in (1.4). Although $L_m(X)$ is an average over an exponentially large number of paths so that at first sight, calculating this quantity may seem practically impossible, there is a recurrence relation which actually gives an algorithm for computing the likelihood ratio in a number of operations which is proportional to the number of nodes, see Section 6 for details. Note that the likelihood ratio $L_m(X)$ is closely related to the partition function of models of random polymers, see [12].

2.1.1 Proof of Theorem 1.2: upper bound

Let h_m be an arbitrary sequence of real numbers tending to infinity and define $\mathcal{S}(h_m)$ as the set of nodes obeying

$$\mathcal{S}(h_m) = \{(i,j) \in V : |j| \le h_m\sqrt{m}\}.$$

In other words, $\mathcal{S}(h_m)$ is a strip of length m and width about $\sqrt{m}$. Our test statistic T_m is just the sum of the variables in the strip,

$$T_m = \sum_{(i,j)\in\mathcal{S}(h_m)} X_{i,j}, \tag{2.2}$$

and we consider a test rejecting for large values of T_m. With the assumption that h_m is going to infinity, the oriented symmetric random walk $(i, S_i)_{0\le i\le m-1}$ is contained in $\mathcal{S}(h_m)$ with large probability. Formally, for each integer a, it follows from the reflection principle that

$$\mathbf{P}\left(\max_{0\le i\le m-1} S_i \ge a\right) \le 2\mathbf{P}(S_{m-1} \ge a).$$

One can then bound the right-hand side by means of the Hoeffding or Azuma's inequality, and obtain

$$\mathbf{P}(S_{m-1} \geq a) \leq e^{-a^2/8(m-1)}.$$

If we then denote by A_m the event $\{\max_{0 \leq i \leq m-1} S_i < h_m \sqrt{m}\}$, it follows from the above two inequalities, that for any sequence (h_m) tending to infinity,

$$\lim_{m \to \infty} \mathbf{P}\left(A_m^c\right) = 0. \tag{2.3}$$

In other words, the random path is contained in the strip $\mathcal{S}(h_m)$ with probability tending to 1.

Now put n_m for the number of nodes in $\mathcal{S}(h_m)$ and note that $n_m = h_m \, m^{3/2}(1 + o(1))$. Under H_0,

$$T_m \sim N(0, n_m)$$

while under H_1, conditionally on the event A_m, we have

$$T_m \sim N(\mu_m, n_m).$$

It then follows from (2.3) that the test which decides H_0 if $|T_m| < m\mu_m/2$ and H_1 otherwise obeys

$$\lim_{m \to \infty} \mathbf{P}_0(\text{Type I}) = 0 \quad \text{and} \quad \lim_{m \to \infty} \mathbf{P}_\pi(\text{Type II}) = 0\,,$$

provided that

$$\mu_m \, m \, n_m^{-1/2} \;\asymp\; \mu_m \, m^{1/4} \, h_m^{-1/2} \to \infty \text{ as } m \to \infty.$$

Since under the assumption of the first part of Theorem 1.2, $\mu_m \, m^{1/4} \to \infty$, one can find a sequence $h_m \to \infty$ with the property $\mu_m \, m^{1/4} \, h_m^{-1/2} \to \infty$. This proves that our test is asymptotically powerful and completes the proof of the first part of Theorem 1.2.

2.1.2 Proof of Theorem 1.2: lower bound

To obtain a lower estimate about the performance of any test, it suffices to bound below the risk of the Bayes test

$$B(\pi) := \inf_{\text{all tests}} \gamma_\pi(T_m) = \mathbf{P}_0(L_m > 1) + \mathbf{P}_\pi(L_m < 1) \tag{2.4}$$

(recall that L_m is the likelihood ratio $L_m(X) = d\mathbf{P}_\pi(X)/d\mathbf{P}_0(X)$). A standard calculation shows that

$$B(\pi) = 1 - \frac{\mathbf{E}_0|L_m - 1|}{2} \geq 1 - \frac{\sqrt{\mathbf{E}_0(L_m - 1)^2}}{2}. \tag{2.5}$$

Therefore, to show that the two hypotheses are asymptotically indistinguishable, it is sufficient to establish that the variance of the likelihood ratio (calculated under the null) tends to zero.

Another standard calculation shows that the variance of L_m is given by

$$\mathbf{E}_0 (L_m - 1)^2 = \mathbf{E}_0 L_m^2 - 1 = \mathbf{E}\, e^{\mu_m^2 N_m} - 1, \tag{2.6}$$

where N_m is the number of crossings of two independent paths drawn from the prior. That is, to derive a lower bound with this strategy, one needs to understand for which sequences (t_m) does the moment generating function $M_m(t_m) := \mathbf{E}\, e^{t_m N_m}$ of the number of crossings of two independent random walks obeys

$$\lim_{m\to\infty} M_m(t_m) = 1. \tag{2.7}$$

When the prior is the distribution of a symmetric random walk, the reader may know that $\mathbf{E}N_m \asymp m^{1/2}$ and since

$$\mathbf{E}e^{t_m N_m} \geq 1 + t_m \mathbf{E}N_m,$$

this shows that it is necessary to have $t_m\, m^{1/2} \to 0$ or equivalently $\mu_m\, m^{1/4} \to 0$. This is the correct asymptotic behavior as we see next.

Let $(S_i)_{1\leq i\leq m}$ and $(S'_i)_{1\leq i\leq m}$ be two symmetric independent random walks (note the slight change of the range of indices which is absolutely unessential). Observe that $\{S_i = S'_i\} = \{S_i - S'_i = 0\}$ so that we equivalently need to study the number N_m of returns to zero of the difference process $(S_i - S'_i)_{1\leq i\leq m}$, which is a Markov chain with the even integers as state space, and with jump probabilities to each neighbor equal to $1/4$, and probability $1/2$ to stay put. Therefore, the joint law of the difference process is that of $(S_{2i})_{1\leq i\leq m}$, where again S is a symmetric random walk (note the doubling of the interval together with the sampling at even times only). An immediate consequence is that

$$\mathbf{P}(N_m = k) = \mathbf{P}\left(|\{1 \leq i \leq m : S_{2i} = 0\}| = k\right).$$

The number of returns of a random walk to the origin has been well studied and we have from [16],[17, Page 96],

$$\mathbf{P}(N_m = k) = \frac{1}{2^{2m-k}} \binom{2m-k}{m}. \tag{2.8}$$

The idea is now to develop a useful upper bound on the right-hand side in order to estimate the moment generating function of N_m.

First, recall the classical refinement of the Stirling approximation to $n!$, see [17, pp. 50-53], which states that

$$\sqrt{2\pi}\, n^{n+1/2}\, e^{-n+\frac{1}{12n+1}} \;<\; n! \;<\; \sqrt{2\pi}\, n^{n+1/2}\, e^{-n+\frac{1}{12n}}.$$

Plugging this approximation in the formula for the binomial coefficient yields

$$\mathbf{P}(N_m = k) \leq \frac{1}{\sqrt{\pi m}} \frac{(1-k/2m)^{2m-k+1/2}}{(1-k/m)^{m-k+1/2}} = \frac{1}{\sqrt{\pi m}} \sqrt{\frac{1-k/2m}{1-k/m}}\, e^{-mg(k/m)}, \tag{2.9}$$

where

$$g(t) = (1-t)\log(1-t) - 2(1-t/2)\log(1-t/2), \quad 0 \le t \le 1.$$

For $t \in (0,1)$, it holds that $d^2/dt^2(g(t) - t^2/4) > 0$ and by convexity, the function $g(t) - t^2/4$ is above its tangent at the origin. This tangent is the line $y = 0$ since $g(0) = g'(0) = 0$, whence

$$g(t) \ge t^2/4, \quad \forall t \in [0,1).$$

Also observe that $(1-t/2)/(1-t) \le 1+t$ for each $t \in [0, 1/2]$. Now fix $0 < \epsilon < 1/2$. For $k \le \epsilon m$, we have $(1-k/2m)/(1-k/m) \le 1+\epsilon$ while for $k < m$ one always has $\sqrt{\frac{1-k/2m}{m(1-k/m)}} \le 1$. We then conclude that

$$\mathbf{P}(N_m = k) \le \begin{cases} \sqrt{\frac{1+\epsilon}{\pi m}}\, e^{-k^2/4m}, & k \le \epsilon m, \\ \frac{1}{\sqrt{\pi}}\, e^{-k^2/4m}, & \epsilon m < k \le m. \end{cases} \tag{2.10}$$

(The case $k = m$ in the above estimate is checked directly rather than from (2.9).)

The estimate (2.10) gives a an upper bound on the moment generating function since

$$M_m(t_m) \le \sum_{k=0}^{\lfloor \epsilon m \rfloor} e^{t_m k} \frac{\sqrt{1+\epsilon}}{\sqrt{\pi m}}\, e^{-k^2/4m} + \sum_{k=\lfloor \epsilon m \rfloor + 1}^{m} e^{t_m k} \frac{1}{\sqrt{\pi}}\, e^{-k^2/4m}.$$

It is clear that if $t_m \to 0$ as $m \to \infty$, the second term of the right-hand side goes to zero as $m \to \infty$, and we focus on the first term. Using the monotonicity in k of both $e^{t_m k}$ and $e^{-k^2 m/4}$, we have

$$\begin{aligned}
\sum_{k=0}^{\lfloor \epsilon m \rfloor} e^{t_m k} \frac{1}{\sqrt{\pi m}}\, e^{-k^2/4m} &\le \frac{1}{\sqrt{\pi m}} + \sqrt{\frac{m}{\pi}} e^{t_m} \int_0^{\epsilon} e^{m\, t_m u}\, e^{-m\, u^2/4}\, du \\
&= \frac{1}{\sqrt{\pi m}} + 2e^{t_m} \int_0^{\epsilon\sqrt{m/2}} e^{\sqrt{2m}\, t_m u}\, \frac{e^{-u^2/2}}{\sqrt{2\pi}}\, du \\
&\le \frac{1}{\sqrt{\pi m}} + 2\, e^{m\, t_m^2 + t_m}\, \mathbf{P}(Z > -\sqrt{2m}\, t_m),
\end{aligned}$$

where Z is a standard normal random variable. It follows that if t_m is chosen such that $\sqrt{m}\, t_m \to 0$ as $m \to \infty$, then

$$\lim_{m\to\infty} 2e^{m\, t_m^2 + t_m}\, \mathbf{P}(Z > -\sqrt{2m}\, t_m) = 1$$

and thus, $\lim_{m\to\infty} M_m(t_m) = 1$. In conclusion, we proved that

$$\mu_m\, m^{1/4} \to 0 \quad \Rightarrow \quad B(\pi) \to 1. \tag{2.11}$$

This proves the second part of Theorem 1.2.

2.2 Minimax detection

Just as in the Bayesian case, we first prove the upper bound by constructing a test which allows us to detect reliably when μ_m decays slower than $(\log m)^{-1/2}$, and then study the lower minimax bound. The idea for obtaining a lower bound is to exhibit a prior on H_1 which makes the Bayesian detection problem as hard as possible. Consider a prior $\bar{\pi}$ on H_1 (here a distribution on the set of paths). Then for all tests T_m,

$$\gamma(T_m) \geq B(\bar{\pi}),$$

where $B(\bar{\pi})$ is the risk of the Bayes test

$$B(\bar{\pi}) = \mathbf{P}_0(L_m > 1) + \mathbf{P}_{\bar{\pi}}(L_m < 1).$$

Our strategy is to construct a prior on the family of paths with a low predictability profile; that is, a process whose location in the future is hard to predict from its current state and history.

2.2.1 Proof of Theorem 1.1: upper bound

Consider a simple test statistic of the form

$$T_m = \sum_{(i,j)\in V} w_{i,j} X_{i,j}, \qquad w_{i,j} := w_i = \frac{\lambda_m}{i+1}. \tag{2.12}$$

Hence, T_m is a weighted sum of the values at the vertices of the graph. For convenience, we fix λ_m so that $\sum_{0\leq i\leq m-1} w_i = 1$. Note that $\lambda_m = (\log m)^{-1}(1+o(1))$. Under H_1, the mean of T_m is given by $\mu_m \sum_{0\leq i\leq m-1} w_i = \mu_m$ and since the $X_{i,j}$'s have identical variance under both H_0 and H_1, we have

$$\mathrm{Var}_0(T_m) = \mathrm{Var}_{1,p}(T_m) = \sum_{(i,j)\in V} w_{i,j}^2 = \sum_{0\leq i\leq m-1} (i+1)w_i^2 = \sum_{0\leq i\leq m-1} \frac{\lambda_m^2}{i+1} = \lambda_m.$$

Hence,

$$T_m \sim_{H_0} N(0,\lambda_m), \quad \text{and} \quad T_m \sim_{H_1} N(\mu_m,\lambda_m)$$

under any alternative. Consider the test which rejects the null whenever $T_m > \mu_m/2$. Then the risk of this test is equal to

$$\gamma(T_m) = 2P\left(N(0,1) > \frac{1}{2}\mu_m\,\lambda_m^{-1/2}\right) \quad \Rightarrow \quad \lim_{m\to\infty} \gamma(T_m) = 0$$

when $\mu_m \lambda_m^{-1/2} \to \infty$ or equivalently when $\mu_m\sqrt{\log m} \to \infty$. This proves the first part of Theorem 1.1.

2.2.2 The predictability profile of a stochastic process

The concept of predictability profile was first introduced in [7].

Definition 2.1. *The predictability profile of a stochastic process* $(S_n)_{n\geq 1}$ *is defined by*

$$\mathrm{PRE}_S(k) = \sup \mathbf{P}(S_{n+k} = x \,|\, S_0, \dots, S_n), \tag{2.13}$$

where the supremum is taken over all positions and histories.

We will consider nearest-neighbor walks where the increments are ± 1. Improving upon earlier results of Benjamini, Pemantle and Peres [7], Häggström and Mossel [19, Theorem 1.4] proved the following:

Theorem 2.2. *Suppose* f_k *is a decreasing positive sequence such that* $\sum_k f_k/k < \infty$. *Then, there exists* $S_0 = 0$ *such that*

$$\mathrm{PRE}_S(k) \leq \frac{C}{kf_k} \tag{2.14}$$

for all $k \geq 1$.

C. Hoffman proved in [20] that this is sharp in the sense that if f_k is a decreasing positive sequence with $\sum f_k/k = \infty$, then the predictability profile (2.14) is impossible to achieve.

In what follows, we will need a quantitative, finite version of Theorem 2.2. This is achieved by using a concrete prior, introduced in [19], which gives the predictability profile below.

Lemma 2.3. *[19, Proposition 3.1] Fix a sequence* $(a_j)_{j\geq 1}$ *obeying* $\sum_j a_j < 1$. *Then there exists a nearest neighbor process* (S_n) *obeying*

$$\mathrm{PRE}_S(k) \leq \frac{20}{ka_{\lfloor \log_2(k/2)\rfloor}}, \qquad \text{for all } k = 1, 2, \dots . \tag{2.15}$$

The construction of the process and the proof of (2.15) may be found in the Appendix. Later on, we will consider a prior on paths obeying (2.15) for suitable values of the sequence (a_j).

2.2.3 Predictability profiles and number of intersections

From now on, we consider stochastic processes with a finite horizon, i.e. $(S_i)_{1\leq i\leq m}$. In the sequel, we will need to estimate the number of times two independent processes drawn from a prior with prescribed predictability profile cross each other. From the proof of [7, Lemma 3.1], we state the following

Lemma 2.4. *Let* B *be such that*

$$\sum_{1\leq k\leq \lfloor m/B\rfloor} \mathrm{PRE}_S(kB) \leq \theta < 1. \tag{2.16}$$

Then for any sequence $(v_n)_{1\leq n\leq m}$ *and all* $k \geq 1$, *the distribution of the number of intersections between* (S_n) *and* (v_n) *obeys*

$$\mathbf{P}(|S \cap v| \geq k) \leq B \cdot \theta^{k/B}, \qquad |S \cap v| := |\{n : S_n = v_n\}|. \tag{2.17}$$

We emphasize that the lemma is valid even if the sequence $(v_n)_{n\geq 1}$ does not determine a nearest neighbor path.

2.2.4 Proof of Theorem 1.1: lower bound

We now prove the lower bound in Theorem 1.1 by providing a lower bound for the Bayes risk $B(\bar{\pi})$ for the prior $\bar{\pi}$ given by Lemma 2.3, and with the m-dependent sequence

$$a_j = a_j(m) := \begin{cases} \frac{\log 2}{3 \log m}, & j \leq \log_2 m\,, \\ 0, & j > \log_2 m\,. \end{cases} \tag{2.18}$$

With the choice above, $\sum a_j \leq 1/3 < 1/2$.

As in the analysis of the Bayes risk, see (2.5),(2.6), we employ the simple bound

$$B(\bar{\pi}) \geq 1 - \frac{\sqrt{\mathbf{E}_0 (L_m - 1)^2}}{2}, \quad \mathbf{E}_0 (L_m - 1)^2 = \mathbf{E} e^{\mu_m^2 N_m} - 1\,, \tag{2.19}$$

where L_m is the likelihood ratio, and N_m is the number of crossings of two independent paths drawn from the prior π. We compute

$$\sum_{k\geq 1} e^{\mu_m^2 k} \mathbf{P}(N_m = k) = \sum_{1\leq k \leq K-1} e^{\mu_m^2 k} \mathbf{P}(N_m = k) + \sum_{k \geq K} e^{\mu_m^2 k} [\mathbf{P}(N_m \geq k) - \mathbf{P}(N_m \geq k+1)],$$

and summing by parts, deduce that

$$\mathbf{E} e^{\mu_m^2 N_m} \leq e^{\mu_m^2 (K-1)} + [1 - e^{-\mu_m^2}] \sum_{k\geq K} \mathbf{P}(N_m \geq k) e^{\mu_m^2 k}.$$

With the choice (2.18), Lemma 2.3 gives

$$\mathrm{PRE}_S(k) \leq \frac{60 \log m}{k \log 2}.$$

In particular, with $B = B_m = 120 (\log m)^2 / \log 2$, we have

$$\sum_{k=1}^{\lfloor m/B_m \rfloor} \mathrm{PRE}_S(k B_m) \leq \frac{1}{2}.$$

Applying Lemma 2.4 yields

$$\begin{aligned} \mathbf{E}_0 L_m^2 &\leq e^{\mu_m^2 (K-1)} + [1 - e^{-\mu_m^2}] B_m \sum_{k \geq K} e^{\mu_m^2 k} \theta^{k/B_m} \\ &\leq e^{\mu_m^2 (K-1)} + [1 - e^{-\mu_m^2}] B_m \frac{a_m^K}{1 - a_m}, \qquad a_m = e^{\mu_m^2} 2^{-1/B_m} < 1, \end{aligned}$$

where the last inequality is due to the fact that $\lim_{m\to\infty} \mu_m^2 B_m = 0$. Further,

$$\liminf_{m\to\infty}(-B_m \log a_m) = \log 2 \quad \Rightarrow \quad \frac{1}{1-a_m} \le \frac{1}{1-e^{-\log 2/(2B_m)}} \le c_1 B_m$$

for some constant c_1 and all m large. It follows that for some constant c_2 and all m large,

$$\mathbf{E}_0 L_m^2 \le e^{\mu_m^2 K} + c_2 \mu_m^2 B_m^2 e^{-K(\log 2)/2B_m}.$$

Taking $K = K_m = 2(B_m \log B_m)/\log 2$ yields, for some constant c_3,

$$\mathbf{E}_0 L_m^2 \le e^{c_3 \mu_m^2 B_m \log B_m} + O(\mu_m^2 B_m) \to 1 \quad \text{as} \quad m \to \infty.$$

Together with (2.19), this concludes the proof of Theorem 1.2.

3 The Complete Binary Tree

In this section, we prove Theorem 1.3. For the upper bound, we show that GLRT is asymptotically powerful if $\mu_m = \mu > \sqrt{2\log 2}$, and that a closely related test is asymptotically powerful if $\mu_m = \mu = \sqrt{2\log 2}$. For the lower bound, we study the likelihood ratio under the uniform prior on paths using a martingale approach.

We start by considering the GLRT, which is based on $M_m = \max\{X_p : p \in \mathcal{P}_m\}$ and X_p is defined in (1.4). We first show that under the null hypothesis, the GLRT obeys

$$\mathbf{P}_0(M_m \ge m\sqrt{2\log 2}) \to 0, \quad m \to \infty.$$

This is in fact a simple application of Boole's inequality and a standard bound on the tail of the normal distribution

$$\mathbf{P}(N(0,1) > t) \le \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}.$$

Indeed,

$$\begin{aligned}\mathbf{P}_0(M_m \ge m\sqrt{2\log 2}) &\le 2^m \mathbf{P}_0(X_p \ge m\sqrt{2\log 2}) \\ &= 2^m \mathbf{P}(N(0,1) \ge \sqrt{2m\log 2}) \le \frac{1}{\sqrt{m\,4\pi\log 2}}.\end{aligned} \tag{3.1}$$

In fact, $M_m/m \to \sqrt{2\log 2}$ a.s., see [27, Section 3]. Under the alternative with $\mu > \sqrt{2\log 2}$, however, the GLRT obeys

$$\mathbf{P}_1(M_m > m\sqrt{2\log 2}) \to 1, \quad m \to \infty. \tag{3.2}$$

Indeed, if p is the path along which the mean is elevated, $M_m \ge X_p$ and X_p/m is normally distributed with mean μ and variance $1/m$.

If $\mu = \sqrt{2\log 2}$, the same argument gives

$$\liminf_{m\to\infty} \mathbf{P}_1(M_m > m\sqrt{2\log 2}) \ge \frac{1}{2}$$

instead of (3.2). This is not quite enough to conclude that H_0 and H_1 can be separated with probability approaching 1. However, taking $m_k = 2^k$, we have from (3.1) and Borel-Cantelli that

$$\mathbf{P}_0(M_{m_k} \geq m_k\sqrt{2\log 2} \text{ infinitely often}) = 0\,,$$

while standard estimates for random walks imply that

$$\mathbf{P}_1(M_{m_k} \geq m_k\sqrt{2\log 2} \text{ infinitely often}) = 1,$$

and even (because the increments $X_p(m_k) - X_p(m_{k-1})$ are exponentially mixing)

$$\liminf_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \mathbf{1}_{\{M_{m_i}\geq m_i\sqrt{2\log 2}\}} \geq \frac{1}{2}\,, \quad \mathbf{P}_1 - \text{ a.s.}$$

Therefore, the test which computes, along the sequence m_k, the number of times $M_{m_k} \geq m_k\sqrt{2\log 2}$ and declares H_0 if this number is less than $k/4$ and H_1 otherwise, has asymptotic full power.

In conclusion, the GLRT or its variant has asymptotic full power if $\mu \geq \sqrt{2\log 2}$.

We now turn to studying the likelihood ratio under the uniform prior π on paths

$$L_m = 2^{-(m-1)} \sum_{\text{all paths } p} e^{\mu X_p - m\mu^2/2},$$

and show that for $\mu < \sqrt{2\log 2}$, its risk

$$B_m(\pi) = \mathbf{P}_0(L_m > 1) + \mathbf{P}_\pi(L_m \leq 1)$$

is bounded away from 0. A lower bound such as (2.5) would not suffice here since we want to recover the same threshold $\sqrt{2\log 2}$. Instead, we turn to martingale methods. Such methods have been used for years, see e.g. [9, 13]. Here we follow the presentation found in [10].

A simple calculation shows that

$$B_m(\pi) = 1 - \mathbf{E}_0(1 - L_m)_+\,.$$

By Proposition 1 in [10], we know that under H_0, L_m is a nonnegative martingale with respect to $\mathcal{F}(X^m)$, where X^m is the set of variables reached by a path in $\mathcal{P}_m$, which converges pointwise to a finite, nonnegative random variable L_∞. Hence, by dominated convergence,

$$\lim_{m\to\infty} B_m(\pi) = 1 - \mathbf{E}_0(1 - L_\infty)_+\,.$$

Applying Proposition 2 in [10], we have that for $\mu < \sqrt{2\log 2}$, L_m is uniformly integrable and therefore $\mathbf{E}_0 L_\infty = 1$. Hence, $\mathbf{P}_0(L_\infty = 0) < 1$ and, consequently,

$$\lim_{m\to\infty} B_m(\pi) > 0\,.$$

We finally briefly argue that if $\mu_m \to 0$, then every sequence of tests is asymptotically powerless. Here, it is enough to use the bound (2.5). It therefore suffices to prove that $\mathrm{Var}_0(L_m) \to 0$ as

$m \to \infty$. Just as before, $\mathrm{Var}_0(L_m) = \mathbf{E}_\pi e^{\mu_m^2 N_m} - 1$ with N_m the number of crossings between two random paths drawn from the prior π. Here, $\mathbf{P}(N_m = k) = 2^{-k}$, $1 \le k \le m-1$, and $\mathbf{P}(N_m = m) = 2^{-m+1}$. In short, the distribution of N_m is that of a truncated geometric random variable with probability of success equal to $1/2$. Set $\tau_m = e^{\mu_m^2}/2$ which is less than 1 for m large. We compute

$$\mathrm{Var}_0(L_m) = \frac{(2\tau_m - 1)(1-\tau_m^m)}{1-\tau_m} \le \frac{2\tau_m - 1}{1-\tau_m}.$$

It is now clear that $\mathrm{Var}_0(L_m) \to 0$ when $\tau_m \to 1/2$ or equivalently when $\mu_m \to 0$. This finishes the proof of the theorem.

4 Extension to Exponential Families

While the previous sections studied the detection problem assuming a Gaussian distribution at the nodes of the graph, it is now time to emphasize that our results hold more generally. In fact, one can obtain similar conclusions for exponential models as well.

Letting F_0 be a distribution on the real line, we define F_θ as the exponential family with associated density $\exp(\theta x - \log\varphi(\theta))$ with respect to F_0 ; note that by definition, $\varphi(\theta) = \mathbf{E}_{F_0}[\exp(\theta X)]$, where $\mathbf{E}_{F_0}$ is the expectation under the distribution F_0. We always assume that $\varphi(\theta) < \infty$ for θ in a neighborhood of 0; further restrictions are mentioned when needed.

Under the null hypothesis, we assume that all the nodes are i.i.d. with distribution F_0 while under $H_{1,m}$, there is a path along which the nodes are i.i.d. with distribution F_{θ_m}, $\theta_m > 0$, and distribution F_0 away from the path. The question is of course for what values of θ_m one can reliably detect this path. To connect this general setup with the previously studied special case, set $\psi(\theta) = \log\varphi(\theta)$ and recall that

$$\mu(\theta) := \mathbf{E}_{F_\theta} X = \psi'(\theta), \quad \text{and} \quad \sigma^2(\theta) := \mathrm{Var}_{F_\theta} X = \psi''(\theta).$$

With these notations, the mean shift is equal to

$$\mu(\theta) - \mu(0) = \psi'(\theta) - \psi'(0) = \psi''(0)(\theta + o(\theta)).$$

In other words, the value of a small mean shift is just about proportional to θ. (In the Gaussian case, $\mu(\theta) = \theta$ and $\log\varphi(\theta) = \theta^2/2$.)

4.1 The regular lattice with an exponential family at the nodes

We consider the minimax detection problem first, and extend Theorem 1.1.

Theorem 4.1. *Suppose that $\theta_m \sqrt{\log m} \to \infty$ as $m \to \infty$. Then there is a sequence of tests which is asymptotically powerful. Conversely, suppose that $\theta_m \log m \sqrt{\log\log m} \to 0$ as $m \to 0$. Then every sequence of tests (T_m) is asymptotically powerless.*

In summary, one can reliably detect a path as long as the mean shift $\mu(\theta_m) - \mu(0) \gg (\log m)^{-1/2}$ while this is impossible if—ignoring the $\sqrt{\log\log m}$ factor—$\mu(\theta_m) - \mu(0) \ll (\log m)^{-1}$.

As an example, consider the case where we have exponentially distributed random variables; under the null, the node variables are exponentially distributed with mean 1 while under the alternative hypothesis, there is a path along which the node variables are exponentially distributed with mean $1+\mu_m$. Call F_0 the density of the exponential with mean 1. The density of an exponential random variable with mean $1+\mu$ with respect to F_0 is given by

$$(1+\mu)^{-1}\, \exp(\mu x/(1+\mu)) := \exp(\theta x - \log\varphi(\theta)),$$

with

$$\theta = \frac{\mu}{1+\mu}, \quad \varphi(\theta) = \frac{1}{1-\theta}.$$

For this exponential model, one can reliably detect a mean shift μ_m if it is significantly larger than $(\log m)^{-1/2}$ while this is impossible if it is much smaller than $(\log m)^{-1}$.

Proof of Theorem 4.1. The proof is similar to that of Theorem 1.1. For the upper bound, we consider the same statistic (2.12) as before, $T_m := \sum_{i,j} w_{i,j} X_{i,j}$, with the exact same choice of weights. Observe first that for any alternative of path p from H_1, the mean difference obeys

$$\mathbf{E}_{1,p}(T_m) - \mathbf{E}_0(T_m) = \mu(\theta_m) - \mu(0).$$

As for the variances, we have

$$\mathrm{Var}_0(T_m) = \sigma^2(0) \sum_{(i,j)\in V} w_{i,j}^2 = \lambda_m \sigma^2(0)$$

and for any alternative with path p from H_1,

$$\begin{aligned}\mathrm{Var}_{1,p}(T_m) &= \sigma^2(0) \sum_{(i,j)\in V} w_{i,j}^2 + [\sigma^2(\theta_m) - \sigma^2(0)] \sum_{0\le i\le m-1} w_i^2 \\ &= \lambda_m \sigma^2(0) + [\sigma^2(\theta_m) - \sigma^2(0)]\, O(\lambda_m^2).\end{aligned}$$

Recall that $\lambda_m = (\log m)^{-1}(1+o(1))$. Using Chebychev's inequality, we see that the probability of Type I and Type II errors go to zero as soon as $[\mu(\theta_m) - \mu(0)]\, \lambda_m^{1/2} \to \infty$ as $m \to \infty$. The first part of the theorem follows from $\mu(\theta_m) - \mu(0) = \theta_m \mathrm{Var}_{F_0}(X)(1+o(1))$. That is, if the mean shift times $\sqrt{\log m}$ increases to infinity, then the probability of each type of error goes to zero.

For the lower bound, we consider the same prior distribution on the family of paths. For exponential models, the variance of the likelihood ratio L_m is given by

$$\mathrm{Var}_0(L_m) = \mathbf{E}[\lambda(\theta_m)^{N_m}] - 1, \qquad \lambda(\theta) = \frac{\varphi(2\theta)}{\varphi(\theta)^2} > 1, \tag{4.1}$$

where again N_m is the number of crossings of two independent paths drawn from the prior; or

$$\mathrm{Var}_0(L_m) = \mathbf{E} e^{\alpha^2(\theta_m)\, N_m} - 1, \qquad \alpha(\theta) = \sqrt{\log\lambda(\theta)}.$$

This is the same expression as before and our previous analysis shows the existence of a prior with the property:

$$\lim_{m\to\infty} \alpha(\theta_m)\, \log m \sqrt{\log\log m} = 0 \quad \Rightarrow \quad \lim_{m\to\infty} \mathrm{Var}_0(L_m) = 0,$$

which implies that the Bayes test is asymptotically powerless. Now it is not difficult to see that for exponential models, $\lambda(\theta) = 1 + O(|\theta|^2)$ so that $\alpha(\theta) = O(\theta)$ for θ close to zero. As a consequence,

$$\lim_{m\to\infty} \theta_m\, \log m \sqrt{\log\log m} = 0 \quad \Rightarrow \quad \lim_{m\to\infty} \alpha(\theta_m)\, \log m \sqrt{\log\log m} = 0,$$

which establishes the second part of the theorem. □

Not surprisingly, the same extension holds in the Bayesian setup as well.

Theorem 4.2. *Consider the uniform prior on paths. Suppose that $\theta_m\, m^{1/4} \to \infty$ as $m \to \infty$, then there is a sequence of tests which is asymptotically powerful. Conversely, if $\theta_m\, m^{1/4} \to 0$ as $m \to 0$, every sequence of tests (T_m) is asymptotically powerless.*

The proof follows that of Theorems 1.2 and 4.1. We omit the details.

4.2 The tree with an exponential family at the nodes

Following [10], define the function f as

$$f(\theta) = \frac{1}{\theta} \log(2\varphi(\theta)). \tag{4.2}$$

By Lemma 4 in [10], f either attains its unique minimum or f is strictly decreasing on $(0, \infty)$. In any case, we denote $\theta^\star \in (0, \infty]$ where f is minimum.

Theorem 4.3. *Assume that $\varphi(\theta) < \infty$ in a neighborhood of $\theta^\star$. If $\theta_m = \theta > \theta^\star$, then the GLRT is asymptotically powerful. If $\theta_m = \theta < \theta^\star$, there does not exist any asymptotically powerful sequence of tests. If $\theta_m \to 0$, then all sequences of tests are powerless. Finally, if $\theta_m = \theta^\star$ then a sequence of asymptotically powerful tests exists.*

For exponential random variables, $\varphi(\theta) = 1/(1-\theta)$ and we numerically compute $\theta^\star \approx .63$. In terms of mean shift (see above), we have $\mu(\theta^\star) - \mu(0) = 1/(1-\theta^\star) - 1 \approx 1.70$. The mean difference along the unknown path must approximately exceed 1.70 to be reliably detectable.

For Bernoulli random variables, $F_\theta = \mathrm{Bernoulli}(e^\theta/(1+e^\theta))$, the function f is decreasing on $(0, \infty)$ and, therefore, $\theta^\star = \infty$. Theorem 4.3 then implies that no asymptotically powerful sequences of tests exist for testing fair coin tossing at the nodes versus biased coin tossing with parameter $q \in (1/2, 1)$ along a path. Note that the situation drastically changes when $q = 1$: in this case, the nodes with value 1 that are connected to the root node through a path of nodes of value 1 form a critical branching process (with an expected number of descendants at each node equal to 1) and, therefore, eventually dies out. Under H_1, however, there is always a path of length m starting from the origin and with all ones. Hence, the test that declares H_1 if one finds such a path and H_0 otherwise is asymptotically powerful.

Proof of Theorem 4.3. The proof is very similar to that of Theorem 1.3. We start with the upper bound, assuming $\theta^\star < \infty$. Define $\xi(t) = \inf_{\theta>0} \varphi(\theta)e^{-t\theta}$. Note that

$$\xi(t) = 1/2 \quad \Leftrightarrow \quad \inf_{\theta>0}(\log(2\varphi(\theta)) - \theta t) = 0 \quad \Leftrightarrow \quad t = \inf_{\theta>0} f(\theta) = f(\theta^\star). \tag{4.3}$$

Because $\varphi(\theta) < \infty$ in a neighborhood of $\theta^\star$, we can replace the estimate (3.1) by the Bahadur-Rao bound [4], which yields

$$\mathbf{P}_0(M_m \geq m\xi^{-1}(1/2)) \leq 2^m \mathbf{P}_0(X_p \geq m\xi^{-1}(1/2)) \leq \frac{C}{\sqrt{m}},$$

for some constant C. (In fact, under our assumptions, $M_m/m \to \xi^{-1}(1/2)$ a.s., by the argument in [27, Section 3].) This estimate and (4.3) imply that

$$\mathbf{P}_0(M_m \geq mf(\theta^\star)) \leq \frac{C}{\sqrt{m}}. \tag{4.4}$$

We now study the behavior of M_m/m under H_1. Let p be the path along which the nodes are sampled from the distribution F_θ. Then the Strong Law of Large Numbers shows that $\lim_{m\to\infty} X_p/m = E_{F_\theta}X$ a.s. and, therefore,

$$\liminf_{m\to\infty} \frac{M_m}{m} \geq \frac{d}{d\theta}(\log \varphi(\theta)) \text{ a.s.}$$

The derivative obeys $d/d\theta(\log\varphi(\theta)) = \varphi'(\theta)/\varphi(\theta) > f(\theta^\star)$ if and only if $\theta > \theta^\star$. This equivalence follows from the identity

$$d/d\theta \, \log(\varphi(\theta)) - f(\theta^\star) = \theta f'(\theta) + f(\theta) - f(\theta^\star).$$

Since f is decreasing on $(0, \theta^\star)$ and strictly increasing on $(\theta^\star, \infty)$, the right-hand side has the sign of $\theta - \theta^\star$. This analysis shows that the GLRT has asymptotic full power if $\theta > \theta^\star$, and the argument handling $\theta = \theta^\star$ is the same as in the Gaussian case, using the full power of (4.4).

The study of the likelihood ratio under the uniform prior over paths is identical to that in the Gaussian case, with the exception that when proving the uniform integrability of the martingale L_m, we use Biggins's theorem (in the form given in cite [24] – noting the condition $\varphi(\theta) < \infty$ in a neighborhood of $\theta^\star$), instead of using Proposition 2 from [10]. (The latter proposition requires that $\varphi(\theta)$ be finite for all $\theta > 0$, or at least for $\theta = 2\theta^\star$.)

5 Extension to Other Graphs

This section emphasizes that results are available for other graphs and in particular, for the analog of the regular lattice in higher dimensions.

- **Regular lattice in dimension** $d' = d + 1$**.** This is the graph with vertex set

  $$V = \{(i, j_1, \ldots, j_d) : 0 \leq i, \; -i \leq j_k \leq i \text{ and } j_k \text{ has the parity of } i\},$$

  and oriented edges $(i, j_1, \ldots, j_d) \to (i+1, j_1 + s_1, \ldots, j_d + s_d)$ where $s_k = \pm 1$.

Consider a distribution from the exponential family at the nodes and the uniform prior on paths. Then the likelihood ratio has been studied in dimension $d+1$—under the name of partition function—in the context of directed random polymers. Martingale methods work well in this context and, the behavior of the likelihood ratio for $d \geq 3$ is similar to the behavior of the likelihood ratio for the tree that we studied in Section 3, see [12, Proposition 3.2.1]. In particular, for $d \geq 3$ there are no asymptotically powerful sequences of tests if $\theta_m = \theta$ obeys $\lambda(\theta)\rho_d < 1$, where $\lambda(\theta)$ is defined in (4.1) and ρ_d is the return probability of a symmetric random walk in dimension d. The results for $d = 2$ only imply that the Bayes risk tends to zero if $\theta_m = \theta > 0$. In contrast, the minimax risk does not go to zero here and this follows from the construction of a prior with low predictability profile. We give a general statement in Theorem 5.3.

To establish a general result, we work with a connected graph (directed or undirected), with one vertex marked that we call origin, and as before we let $\mathcal{P}$ be the set of self-avoiding paths starting at the origin and $\mathcal{P}_m \subset \mathcal{P}$ be the subset of paths of length m. Under the null hypothesis, all the nodes are i.i.d. F_0 while under the alternative, there is a path in $\mathcal{P}_m$ along which the nodes are i.i.d. F_1. We assume throughout that F_1 is absolutely continuous with respect to F_0, for otherwise the detection problem becomes trivial.

Definition 5.1. *A distribution π on $\mathcal{P}$ is said to have an exponential intersection tail with parameter $\eta \in (0,1)$ if there exists $C > 0$ such that if N is the number of crossings of two independent samples from π,*

$$\mathbf{P}(N \geq k) \leq C \cdot \eta^k, \quad \forall k \geq 1.$$

The regular lattice with $d \geq 2$ (i.e., $d' \geq 3$) admits a measure on paths with an exponential intersection tail [7, Theorem 1.3]. Note that a summable predictability profile implies an exponential intersection tail.

Definition 5.2. *Let $L = dF_1/dF_0$ be the likelihood ratio for testing F_1 versus F_0 at a singe node. The Pearson χ^2-distance between F_0 and F_1 is defined as $\chi^2(F_0, F_1) = \mathrm{Var}_0(L)$.*

With these definitions, we have the following general statement:

Theorem 5.3. *Suppose that there is a distribution π on $\mathcal{P}$ having an exponential intersection tail with parameter η. Then if $\chi^2(F_0, F_1) < \eta^{-1} - 1$, there are no asymptotically powerful sequences of tests.*

The proof does not require any argument that we have not already presented, and is omitted. For exponential variables, $\chi^2(F_0, F_\theta) = \lambda(\theta) - 1$, where $\lambda(\theta)$ is defined in (4.1) and, therefore, no asymptotically powerful sequences of tests exist if $\lambda(\theta)\eta < 1$.

Theorem 5.3 provides a lower bound on the minimax threshold for reliable detection. For an upper bound, suppose for example that the variables are exponentially distributed and assume that $\#\mathcal{P}_m = O(\delta^m)$ for some positive constant δ; for instance, $\delta = 2^d$ works for the regular lattice in dimension $d+1$. An application of Boole's inequality and of the Law of Large Numbers shows that, under those assumptions, the GLRT is asymptotically powerful if $\xi(\theta)\delta > 1$, where again $\xi(t) = \inf_{\theta > 0} \varphi(\theta) e^{-t\theta}$.

6 Numerical Experiments

We now explore the empirical performance of some of the detection methods we proposed for the regular lattice. The variables at the nodes are independent Gaussians. To measure the performance, we fix the probability of Type I error at 5% and estimate the power or detection rate, i.e. the probability of deciding in favor of the alternative H_1 when H_1 is true. This power function was estimated at values of the mean shifts μ (the mean of the node variables along the path) at which this function is varying.

6.1 Bayesian detection under the uniform prior

We first consider detection under the uniform prior on paths. We compare the performance of the Bayes test, the GLRT and the test based on the Strip statistic which was used in the proof of the upper bound in Theorem 1.2. The Bayes test is optimal in this setting and we recall that the Strip statistic was shown to achieve the optimal detection rate. This paper did not theoretically analyze the performance of the GLRT in this situation, however, and we would like to do so empirically.

6.1.1 Computing the Bayes statistic

As emphasized earlier, there exists a rapid algorithm for calculating the Bayes statistic $L_m(X)$ (2.1). Consider any node $v = (i, j)$ ($0 \le i \le m-1$ and j has the parity of i) and let $\mathcal{P}^{\text{End}}(v)$ be the set of paths starting at the root $(0, 0)$ and ending at the node v. Set

$$Y(v) := 2^{-i} \sum_{p \in \mathcal{P}^{\text{End}}(v)} e^{\mu X_p - (i+1)\mu^2/2}.$$

With these notations, $L_m(X)$ is the sum of Y over all the terminal nodes v for which $i = m-1$. Now observe the recurrence

$$Y(v) = e^{\mu X_v - \mu^2/2} \cdot \frac{Y(v^+) + Y(v^-)}{2}, \tag{6.1}$$

where (v^+, v^-) are the two predecessors of v in the graph, i.e. the two nodes from which one can reach v in one step. (By convention, set $Y(v^\pm) = 0$ if $v^\pm$ is outside of the triangle.) This recurrence shows that one can compute the Bayes statistics in $O(m^2)$ flops.

For each value of μ and m then, we simulated the Bayes statistic under H_0 and H_1 using 2,000 realizations for each. Here and below, each realization uses a new path realization drawn from the uniform distribution.

6.1.2 Simulating the Strip statistic

For a positive integer B, the strip statistic $T_{m,B}$ is the sum of the random variables falling in the centered strip of length m and width $2B + 1$,

$$T_{m,B} = \sum_{0 \le i \le m-1} \; \sum_{j : |j| \le \min(i, B)} X_{i,j}.$$

m	1,025	2,049	4,097	8,193	16,385
$\mu_{.95}$ (Bayes)	0.37	0.31	0.26		
$\mu_{.95}$ (Strip)	0.84	0.69	0.59	0.51	0.42
$\mu_{.95}$ (GLRT)	0.46	0.40	0.36	0.33	

Table 1: Value of the mean shift giving a detection rate of about 95% when using the Bayes test, the Strip statistic test with width $B = 2\sqrt{m}$ and the GLRT (uniform prior on paths). One can compute $\mu_{.95}$ for the Strip statistic for large values of m since it is given analytically.

Under H_0, $T_{m,B} \sim N(0, n_{m,B})$ where $n_{m,B}$ is the number of vertices in the strip while under H_1, $T_{m,B} \sim N(\mu \cdot R_{m,B}, n_{m,B})$ where $R_{m,B}$ is the number of vertices inside the strip that the random path visits. Therefore, one can simulate $T_{m,B}$ by taking one realization of $R_{m,B}$ multiplying it by μ and adding an independent mean-zero Gaussian variable.

It remains to choose the width of the strip. We ran simulations with $B = \nu\sqrt{m}$ for $\nu = 0.75, 1, 2, 3$. Among these values, $B = 2\sqrt{m}$ gave the best performance (at least for the graph sizes we considered). Finally, for a fixed μ and m, we used 5,000 realizations of the test statistic to estimate the detection rate.

6.1.3 Simulating the GLRT

The GLRT statistic rejects for large values of $M_m = \max\{X_p : p \in \mathcal{P}_m\}$. This statistic can be calculated rapidly using dynamic programming; e.g. Dijkstras algorithm [1] has here a computational complexity proportional to the number of nodes. For each graph size, the threshold corresponding to a Type I error probability approximately equal to .05 and the detection rate for a fixed μ were based on 10,000 and 1,000 realizations respectively.

6.1.4 Comparing the tests

To compare the three tests, one can estimate the value of the mean shift which gives a detection rate of about 95% from graphs plotting the detection rates versus μ (see Figure 6). Call this quantity $\mu_{.95}$. Table 1 shows $\mu_{.95}$ for the Bayes test, the test based on the Strip statistic test and the GLRT for different graph sizes. As expected, the Bayes test outperforms the other two but one needs to recall that those tests do not require information about the parameter μ while the Bayes test does. Figure 3 shows a log-log plot of $\mu_{.95}$ as a function of m together with least-squares line fits. The slope of the line is -0.255 for the Bayes test and -0.246 for the Strip test. Both of these values are quite close to the $-1/4$ exponent one finds in Theorem 1.2. For the GLRT, the slope is about -0.16. This suggests that the Strip statistic test might eventually outperform the GLRT for sufficiently large graphs. The fitted lines meet at approximately $m = 2^{20} \approx 10^6$ but it would be computationally extremely intensive to run simulations for graphs of this size. The point here is that these simulations suggest that the GLRT is only able to detect at $\mu \approx m^{-1/6}$ and, therefore, does not achieve the optimal detection rate under the uniform prior on paths.

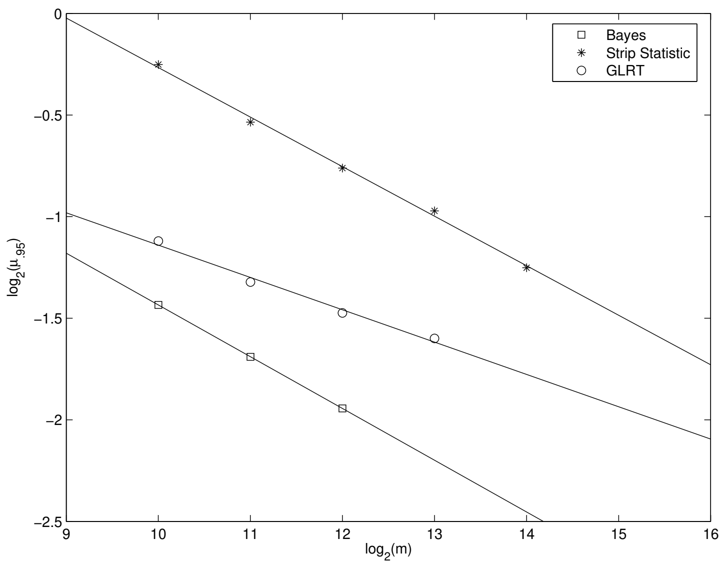

Figure 3: Comparison of the Bayes test, the Strip statistic test and the GLRT under the uniform prior. The plot shows the value $\mu_{.95}$ of the mean shift for which a given test achieves a 95% detection rate when the rate of false alarm is set at 5% as a function of the graph size m (log-log scale).

m	1,025	2,049	4,097	8,193	16,385	32,769
$\mu_{.95}$ (WAS)	1.20	1.15	1.10	1.06	1.03	0.99
$\mu_{.95}$ (GLRT)	0.90	0.89	0.885	0.88		

Table 2: Value of the mean shift giving a detection rate of about 95% when using the WAS test and the GLRT for detecting the increasing path. One can compute $\mu_{.95}$ for the WAS for large values of m since it is given analytically.

6.2 Minimax detection

We focus here on the increasing path p, where $p_i = i, 0 \leq i \leq m-1$, as we believe this path to be the most challenging for the GLRT. In this section, we compare the performance of the GLRT with The Weighted Average Statistic test (WAS) defined in (2.12).

Recall that the WAS is distributed as $N(0, \lambda_m)$ under H_0 and as $N(\mu, \lambda_m)$ under H_1 regardless of the unknown path ($\lambda_m \sim (\log m)^{-1}$). Thus to achieve a power equal to .95 at the 5% significance level, we need $\mu \geq 2z_{.95}\sqrt{\lambda_m}$, where $z_{.95}$ is the 95% standard normal quantile. Some power curves for the WAS are graphed in Figure 5. We use simulations to graph similar curves for the GLRT, see Figure 6; each point is based on 1,000 realizations of the statistic.

While the power curves for the WAS tend to translate to the left, this does not seem to be the case for the GLRT. This might indicate that the detection threshold for the GLRT does not tend to zero as m increases, just as in the case of the binary tree.

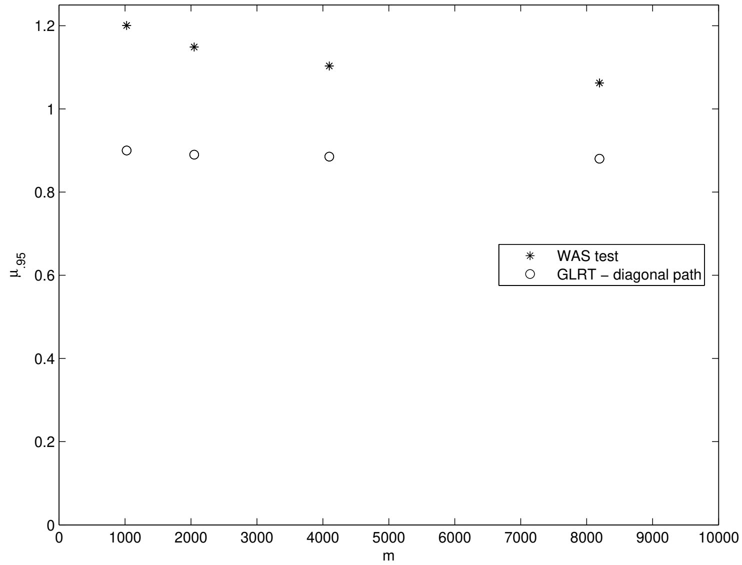

Figure 4: Comparison of the GLRT and the WAS when the anomalous path is the increasing path. The plot shows the value $\mu_{.95}$ of the mean shift for which a given test achieves a 95% detection rate when the rate of false alarm is set at 5% as a function of the graph size m

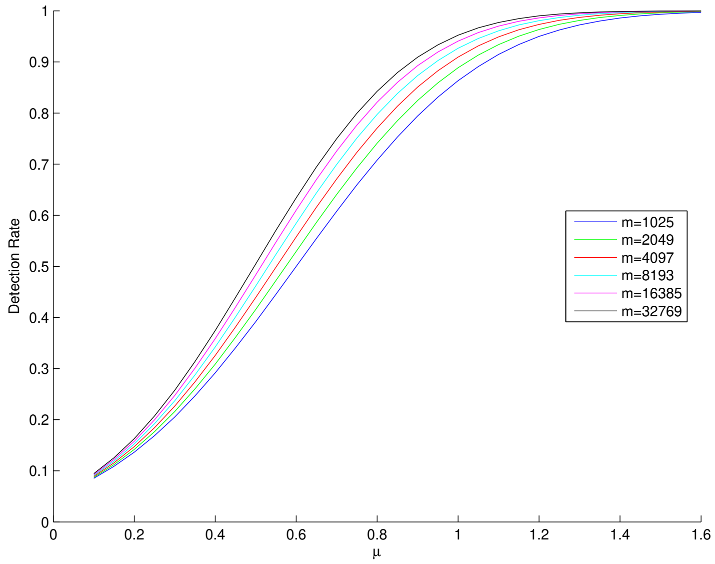

Figure 5: Detection rate curves for the WAS statistic with $m = 1025, 2049, 4097, 8193, 16385, 32769$. As m increases, the curve moves to the left. The Type I error is set to 5%.

7 Discussion

Our paper leaves a number of open questions, invites several refinements, and we briefly discuss some of these.

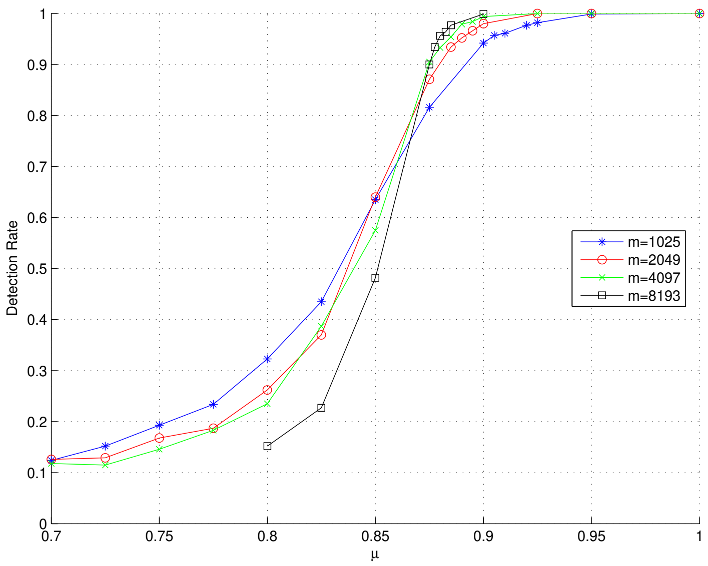

Figure 6: Detection rate curves for the GLRT (increasing path). The probability of Type I error is set to 5%.

7.1 Sharpening the minimax detectability threshold in the two dimensional regular lattice

There is a gap between the upper and lower bounds in Theorem 1.1: the detection threshold for our estimator (2.12) is of order $\mu_m \sim (\log m)^{-1/2}$, but the priors we constructed showed non-detectability only when $\mu_m \sim (\log m)^{-1}$ (ignoring loglog factors). We do not see how to improve our prior to yield significantly better bounds, and it seems that in any case, explicit priors of this family—as constructed in [19] for example—will not yield a lower bound obeying $\mu_m \gg (\log m)^{-3/4}$. It would be very interesting to understand this better and decide what is the actual rate of the detectability threshold. In a personal communication, C. Hoffman informed the authors that he was able to close the gap in Theorem 1.1.

With this in mind, we would like to emphasize that the test (2.12) used to prove the upper bound in Theorem 1.1 does not use the "continuity" of the path, only that it is known to be in the triangle. That is, the test detects any sequence of the form $\{(i, p_i) : 0 \leq i \leq m-1\}$ as long as (i, p_i) is a vertex in the graph provided, of course, that μ_m is of order $(\log m)^{-1/2}$. In fact, $(\log m)^{-1/2}$ turns out to be the minimax detection threshold when the set of vertices with positive mean is any sequence (i, p_i) remaining in the triangle. Indeed, the least favorable prior chooses the (p_i) independently and uniformly at random in their respective range so that the number of crossings of two independent paths obeys

$$N_m = \sum_{1 \leq i \leq m} I_i,$$

where the I_i's are independent with $\mathbf{P}(I_i = 1) = 1/i$ and $\mathbf{P}(I_i = 0) = 1 - 1/i$. The same argument as before shows that

$$\mathbf{E}_0(L_m - 1)^2 = \mathbf{E}e^{\mu_m^2 N_m} - 1 = \prod_{1 \le i \le m} \left(1 + \frac{e^{\mu_m^2} - 1}{i}\right) - 1,$$

which is easily shown to converge to zero when $\mu_m (\log m)^{1/2} \to 0$.

7.2 Studying the GLRT on the two dimensional regular lattice

The GLRT may not be (near) optimal at all in the minimax sense. A strong indication of that can be deduced from work of Baik and Rains in [5, Section 4.4] and [6]. In the language of the current paper, they deal with the following problem: consider directed paths in the triangle $\{(i,j) \in \mathbf{Z}^2 : j \le i \le m\}$. That is, starting from the origin $(0,0)$, a path is a sequence of increments by 1 unit in the right or upward direction (this corresponds to a rotation of the regular graph considered in Theorem 1.1 with its lower half erased). Under H_0, all vertices are i.i.d. exponential random variables with parameter 1. Under H_1, the variables along the "diagonal path" (the path $(0,0)$, $(1,1)$, $(2,2)$ and so on) are i.i.d. exponential with mean $1+\mu$ (of course, in this situation, H_1 is asymptotically distinguishable from H_0 if $\mu > 0$ but this is of no concern in what follows). They consider the GLRT statistic M_m, which consists of the maximum partial sums among all possible directed paths connecting $(0,0)$ to (m,m), and show that the limit distribution of (a properly rescaled version of M_m) does not depend on μ as long as $\mu < 1$ (this follows from the geometric case treated in [5, Section 4.4]). This strongly hints that in this particular setup, the GLRT is far from optimal since Section 2 shows that the minimax risk with respect to all possible directed path goes to zero for any $\mu > 0$. (Note that strictly speaking, since the mode of convergence in [5] is weak convergence and not total variation, the results there hint, but do not imply, that the GLRT is not optimal.) Our simulations in Section 6.2 provide additional evidence for Gaussian variables.

Also of interest would be to study the power of the GLRT with a uniform prior on paths where we suspect that the GLRT does not achieve the optimal threshold.

7.3 Unknown starting location

Throughout this paper, we assumed that under H_1, the unknown path starts at a known node (the origin). The same question can be posed also when the starting location is not known. For concreteness, consider the regular lattice as in Section 2, and allow the unknown path of length $m/2$ to start at any vertex in the collection $\{(i,j)\}_{i=0}^{m/2}$. Does there exist an asymptotically powerful test (in the minimax sense) for some sequence $\mu_m \to 0$? Similarly, we could also imagine having a square lattice $V = (i,j)$ with $0 \le i \le m-1$, $0 \le j < 2m$ (j has the parity of i as before), and with edges $(i,j) \to (i+1, j+s)$ where $s = \pm 1$ and $j+s$ is understood modulo $2m$. If we know the starting location $(0,j)$ of the unknown path of length m, then this is the model problem discussed in Section 2. But studying this problem when we do not know the starting vertex is also of interest.

7.4 Further refinements

In this paper, we assumed that the node variables are independent and identically distributed and clearly, one could address similar testing problems in far more general setups. Interesting extensions include situations in which the variables are correlated or in which the means along the unknown path are not all equal. Following up on the nonparametric signal detection problem, one could also imagine problems where the vector of means is not exactly sparse in the sense that it is zero away from the unknown path but only rapidly decaying away from this path.

While this paper focuses on asymptotic properties of the detection problem, it is also of interest to develop test statistics with good finite sample size properties, and we hope to report on our progress in a future publication.

7.5 Other work

While this paper was being written, N. Berger and Y. Peres described to us some of their own results, obtained independently, which address related problems and answer some of the questions raised above.

Acknowledgments. Emmanuel Candes and Ofer Zeitouni would like to thank Houman Ohwadi for fruitful conversations. Ofer Zeitouni would like to thank C. Hofman and Y. Peres for describing some of their unpublished results to him, and J. Baik for explaining to him the relevance of [5], see the discussion in Section 7.

8 Appendix: Proof of Lemma 2.3

To construct a stochastic process obeying (2.15), we follow [19] and let S_n be the sum $S_n = \sum_{i=1}^n I_i$ with $\mathbf{P}(I_i = 1) = p_i$ and $\mathbf{P}(I_i = -1) = 1 - p_i$. Here the p_i's are stochastic (random environment) and defined by

$$p_i = 1/2 + p_i^{(1)} + p_i^{(2)} + \dots ,$$

where $(p_i^{(1)})$, $(p_i^{(2)})$, $\dots$, are independent processes.

1. For each i and j, the distribution of $p_i^{(j)}$ is uniform on $[-a_j, a_j]$.
2. The value $p_i^{(j)}$ is constant in i for $i = 1, \dots , 2^j$. At time $2^j + 1$ it switches to a new independent value uniform on $[-a_j, a_j]$ which is kept until time 2×2^j and so on.

Note that we need

$$\sum_j a_j < 1/2 \tag{8.1}$$

for this to make sense so that the $p_i \in (0, 1)$. Finally the I_i's are independent conditioned on the random environment (p_i).

With this in place, Häggström and Mossel in [19, Proposition 3.1] showed that there exists a nearest neighbor process (S_n) obeying

$$PRE_S(k) \leq \frac{C}{ka_{\lfloor \log_2(k/2) \rfloor}}, \qquad \text{for all } k = 1, 2, \dots . \tag{8.2}$$

where $C = 4[C_1 + 1]$, with $C_1 = 2^{m_k} a_{m_k} \cdot \mathbf{P}(Y < \mathbf{E}Y/2)$, $m_k = \lfloor \log_2(k/2) \rfloor$ and Y is a binomial random variable with 2^{m_k} trials and a probability of success equal to a_{m_k}. Since for any binomial variable $Y_{n,p} \sim \mathrm{Bin}(n, p)$,

$$np\,\mathbf{P}(Y_{n,p} < np/2) \leq \frac{4np\mathrm{Var}(Y_{n.p})}{n^2p^2} \leq 4,$$

$C_1 \leq 4$ and thus the constant $C \leq 20$.

As discussed earlier, this remark is of importance to us since we have used a sequence (a_j) that depends explicitly on m.

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows. Theory, algorithms and applications.* Prentice-Hall, New-York, 1993.

[2] E. Arias-Castro, D. Donoho, and X. Huo. Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory*, 51(7):2402–2425, 2005.

[3] E. Arias-Castro, D. Donoho, and X. Huo. Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.*, 34(1), 2006.

[4] R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Ann. Math. Statis.*, 31:1015–1027, 1960.

[5] J. Baik and E. M. Rains. The asymptotics of monotone subsequences of involutions. *Duke Math. J.*, 109(2):205–281, 2001.

[6] J. Baik and E. M. Rains. Symmetrized random permutations. In *Random matrix models and their applications*, volume 40 of *Math. Sci. Res. Inst. Publ.*, pages 1–19. Cambridge Univ. Press, Cambridge, 2001.

[7] I. Benjamini, R. Pemantle, and Y. Peres. Unpredictable paths and percolation. *Ann. Probab.*, 26(3):1198–1211, 1998.

[8] S. Bhavsar and R. Splinter. The superiority of the minimal spanning tree in percolation analyses of cosmological data sets. *Monthly Notices of the Royal Astronomical Society*, 282:1461–1466, 1996.

[9] J. D. Biggins. Martingale convergence in the branching random walk. *J. Appl. Probability*, 14(1):25–37, 1977.

[10] E. Buffet, A. Patrick, and J. V. Pulé. Directed polymers on trees: a martingale approach. *J. Phys. A*, 26(8):1823–1834, 1993.

[11] E. J. Candes, P. R. Charlton, and H. Helgason. Detecting highly oscillatory signals by chirplet path pursuit, 2006. Technical report, California Institute of Technology.

[12] F. Comets, T. Shiga, and N. Yoshida. Probabilistic analysis of directed polymers in random environments: a review. In *Stochastic analysis on large scale interacting systems*, pages 115–142, Tokyo, Japan, 2004. Math. Soc. Japan.

[13] B. Derrida and H. Spohn. Polymers on disordered trees, spin glasses, and traveling waves. *J. Statist. Phys.*, 51(5-6):817–840, 1988. New directions in statistical mechanics (Santa Barbara, CA, 1987).

[14] D. L. Donoho and X. Huo. Beamlets and multiscale image analysis. In *Multiscale and multiresolution methods*, volume 20 of *Lect. Notes Comput. Sci. Eng.*, pages 149–196. Springer, Berlin, 2002.

[15] B. Efron and R. Tibshirani. On testing the significance of sets of genes, 2006. Technical report, Stanford University.

[16] W. Feller. The numbers of zeros and of changes of sign in a symmetric random walk. *Enseignement Math. (2)*, 3:229–235, 1957.

[17] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[18] J. Glaz, J. Naus, and S. Wallenstein. *Scan statistics.* Springer Series in Statistics. Springer-Verlag, New York, 2001.

[19] O. Häggström and E. Mossel. Nearest-neighbor walks with low predictability profile and percolation in $2+\epsilon$ dimensions. *Ann. Probab.*, 26(3):1212–1231, 1998.

[20] C. Hoffman. Unpredictable nearest neighbor processes. *Ann. Probab.*, 26(4):1781–1787, 1998.

[21] Y. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics.* Springer-Verlag, New York, 2003.

[22] R. R. Kompella, S. Singh, and G. Varghese. On scalable attack detection in the network. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 187–200, New York, NY, USA, 2004. ACM Press.

[23] M. Kulldorff. A spatial scan statistic. *Comm. Statist. Theory Methods*, 26(6):1481–1496, 1997.

[24] R. Lyons. A simple path to Biggins' martingale convergence for branching random walk. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 217–221. Springer, New York, 1997.

[25] V. Martinez and E. Saar. *Statistics of the Galaxy Distribution.* Chapman and Hall/CRC press, Boca Raton, 2002.

[26] G. P. Patil, J. Balbus, G. Biging, J. Jaja, W. L. Myers, and C. Taillie. Multiscale advanced raster map analysis system: definition, design and development. *Environ. Ecol. Stat.*, 11(2):113–138, 2004.

[27] R. Pemantle. Tree-indexed processes. *Statist. Sci.*, 10(2):200–213, 1995.

[28] J.-L. Starck, V. Martinez, D. Donoho, O. Levi, P. Querre, and E. Saar. Analysis of the spatial distribution of galaxies by multiscale methods. *EURASIP Journal on Applied Signal Processing*, 15:2455–2469, 2005.

[29] Y. Zhong, A. Jain, and M.-P. Dubuisson-Jolly. Object tracking using deformable templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2(5):544–549, 2000.

[30] B. Zupan, J. Demsar, I. Bratko, P. Juvan, J. A. Halter, A. Kuspa, and G. Shaulsky. Genepath: a system for automated construction of genetic networks from mutant data. *Bioinformatics*, 19(3):383–389, 2003.