

Scaling limits for the threshold window: When does a monotone Boolean function flip its outcome?

Daniel Ahlberg^a and Jeffrey E. Steif^b

With an appendix by Gábor Pete^c

^a*Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brasil, Department of Mathematics, Uppsala University, SE-75106 Uppsala, Sweden. E-mail: ahlberg@impa.br; url: www.impa.br/~ahlberg*

^b*Mathematical Sciences, Chalmers University of Technology, Mathematical Sciences, University of Gothenburg, SE-41296 Gothenburg, Sweden. E-mail: steif@chalmers.se; url: http://www.math.chalmers.se/~steif*

^c*Rényi Institute, Hungarian Academy of Sciences, Budapest, 13-15 Reáltanoda u., 1053 Budapest, Hungary, Institute of Mathematics, Budapest University of Technology and Economics, 1 Egry József u., 1111 Budapest, Hungary. E-mail: gabor@math.bme.hu; url: http://www.math.bme.hu/~gabor*

Received 24 August 2015; revised 11 August 2016; accepted 12 August 2016

Abstract. Consider a monotone Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ and the canonical monotone coupling $\{\eta_p : p \in [0, 1]\}$ of an element in $\{0, 1\}^n$ chosen according to product measure with intensity $p \in [0, 1]$. The random point $p \in [0, 1]$ where $f(\eta_p)$ flips from 0 to 1 is often concentrated near a particular point, thus exhibiting a threshold phenomenon. For a sequence of such Boolean functions, we peer closely into this threshold window and consider, for large n , the limiting distribution (properly normalized to be nondegenerate) of this random point where the Boolean function switches from being 0 to 1. We determine this distribution for a number of the Boolean functions which are typically studied and pay particular attention to the functions corresponding to iterated majority and percolation crossings. It turns out that these limiting distributions have quite varying behavior. In fact, we show that any nondegenerate probability measure on \mathbb{R} arises in this way for some sequence of Boolean functions.

Résumé. Soit $f : \{0, 1\}^n \rightarrow \{0, 1\}$ une fonction booléenne monotone, et $\{\eta_p : p \in [0, 1]\}$ le couplage monotone canonique d'éléments de $\{0, 1\}^n$ choisis selon la mesure produit d'intensité $p \in [0, 1]$. Le point aléatoire $p \in [0, 1]$ en lequel $f(\eta_p)$ bascule de 0 à 1 est souvent concentré près d'une valeur particulière, présentant ainsi un effet de seuil. Pour une suite de telles fonctions booléennes, nous étudions de plus près la fenêtre de seuil correspondante en considérant la loi limite lorsque n tend vers l'infini (proprement normalisée pour être non-dégénérée) de ce point aléatoire critique où la fonction booléenne bascule. Nous déterminons cette loi pour de nombreuses fonctions booléennes classiques, en portant une attention particulière aux cas de la majorité itérée et des croisements de percolation. Il se trouve que ces lois limites ont des comportements d'une grande variété : en fait, nous montrons que toute mesure de probabilité non-dégénérée sur \mathbb{R} peut être obtenue de cette façon à partir d'une suite bien choisie de fonctions booléennes.

MSC: 06E30; 60F20; 60F99; 60K35

Keywords: Boolean functions; Sharp thresholds; Influences; Iterated majority function; Near-critical percolation

1. Introduction

It has been known for quite some time that typical events involving many independent random variables exhibit “thresholds” in the sense that the probability of the given event changes sharply as the parameter of the independent random variables varies. Observations of this kind were first made in the context of random graphs by Erdős and Rényi

[11]. A more general understanding of the existence of threshold phenomena has since then been obtained through a series of papers. For instance, Russo [47] showed that a monotone event defined in terms of a family of independent Bernoulli variables exhibits a threshold if its dependence on each variable is small. Russo’s result was later refined by Talagrand [52]. The first estimates on the “sharpness” of the threshold were obtained by Friedgut and Kalai [14], critically building on work originating from Kahn, Kalai and Linial [31]. Related results also appeared in [4,13] and elsewhere; see also [32] for a more extensive overview of the field.

Less is known when it comes to closer inspections of the “threshold window”. Although the windows corresponding to certain graph properties are well understood, there is to our knowledge no general study of this transition. We aim with the present paper to offer a unified perspective on threshold transitions, and show that these transitions present quite varying behavior.

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a monotone (increasing) Boolean function and assign $[0, 1]$ -uniform random variables $\xi_1, \xi_2, \dots, \xi_n$ to the elements of $[n] := \{1, 2, \dots, n\}$. For $p \in [0, 1]$, let $\eta_p = \{i \in [n] : \xi_i \leq p\}$, and note that

- (a) each $i \in [n]$ is present in η_p with probability p independently for different i ; and
- (b) $i \in \eta_p$ implies $i \in \eta_{p'}$ for $p' > p$.

Thus, η_p corresponds to an element $\omega \in \{0, 1\}^n$ chosen according to product measure with intensity p , in the sequel denoted by \mathbb{P}_p , and $(\eta_p)_{p \in [0,1]}$ constitutes the standard monotone coupling of elements in $\{0, 1\}^n$ chosen according to \mathbb{P}_p , as p varies between 0 and 1. We study the random point p at which $f(\eta_p)$ changes from 0 to 1. For a given sequence of monotone Boolean functions $(f_n)_{n \geq 1}$, our goal will be to find the (nondegenerate) limiting distribution of this random point after proper normalization, should it exist. In the most interesting examples, one has a threshold phenomenon where, for large n , $\mathbb{P}_p(f_n)$ goes from 0 to 1 within a very small interval, which results in this transition point having a degenerate limit; one then needs to renormalize in order to obtain a nondegenerate limiting distribution. (If there is no threshold phenomenon, then this point should already have a nondegenerate limit and no further analysis is made.) One of our goals is to describe the distribution of this random point for some commonly studied Boolean functions.

To be more precise, given a monotone Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, we define the random variable

$$T(f) := \min\{p \in [0, 1] : f(\eta_p) = 1\};$$

this is the point where f switches from 0 to 1 in the canonical coupling. Given a sequence $(f_n)_{n \geq 1}$ of monotone Boolean functions $f_n : \{0, 1\}^n \rightarrow \{0, 1\}$, we want to find, if possible, normalizing constants $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ with the a_n ’s nonnegative such that $a_n(T(f_n) - b_n)$ converges, as $n \rightarrow \infty$, to a nondegenerate limiting distribution, and to determine what that limit may be. Observe that for $x \in \mathbb{R}$

$$\mathbb{P}(a_n(T(f_n) - b_n) \leq x) = \mathbb{P}(T(f_n) \leq b_n + x/a_n) = \mathbb{P}_{p_n}(f_n(\omega) = 1), \tag{1}$$

where $p_n = b_n + x/a_n$. Recall the theorem of types (see [9, Theorem 3.7.5]) which tells us that there is *essentially* only one way to normalize a sequence of random variables and that there is *essentially* at most one possible nondegenerate limiting distribution. (The word “essentially” in the latter part of the statement means “up to a change of variables of the form $x \mapsto ax + b$ ”).

Notation. When understood from the context, we will write T_n for $T(f_n)$.

The dictatorship function, which for every n outputs the value of the first coordinate of the input, clearly has that T_n is uniformly distributed on the interval $[0, 1]$ for each n and so no scaling is needed. A simple example where scaling is needed is the OR function which is 1 if and only if at least one bit is 1. In this case, it is immediate to check that nT_n converges in distribution to a unit exponential random variable. The cases when nontrivial normalization is needed are exactly those covered in the next definition.

Definition. We say that $(f_n)_{n \geq 1}$ has a *threshold* if there exists a sequence $(p_n)_{n \geq 1}$ such that for all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{p_n + \varepsilon}(f_n = 1) = 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{P}_{p_n - \varepsilon}(f_n = 1) = 0.$$

This is equivalent to $T_n - p_n$ approaching 0 in distribution. (Often p_n will not depend on n .)

We give an alternative description of $T(f)$ which will be useful to have in mind, in particular when we study the Boolean function known as “tribes”. Recall that a 1-witness for f is a minimal set $W \subseteq [n]$ such that $\{\omega_i = 1 \text{ for all } i \in W\}$ implies $f(\omega) = 1$. Similarly, a 0-witness for f is a minimal set $W \subseteq [n]$ such that $\{\omega_i = 0 \text{ for all } i \in W\}$ implies $f(\omega) = 0$. Witnesses may be used to characterize $T(f)$. Writing \mathcal{W}^1 for the set of 1-witnesses and \mathcal{W}^0 for the set of 0-witnesses for f , it is immediate to check that for any monotone Boolean function one has

$$T(f) = \min_{W \in \mathcal{W}^1} \max_{i \in W} \xi_i = \max_{W \in \mathcal{W}^0} \min_{i \in W} \xi_i. \tag{2}$$

We next briefly discuss what one should *expect* the normalizing constants $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ to be in typical situations. Certainly it is reasonable that b_n should be close to $\mathbb{E}[T_n]$. In cases where we have a threshold, heuristically, the size of the “threshold interval” around p_n where $\mathbb{P}_p(f_n = 1)$ moves from being near 0 to being near 1 *should* be governed by $\frac{d}{dp} \mathbb{P}_p(f_n = 1)$ evaluated at $p = p_n$. The Margulis-Russo formula (see e.g. [17, Theorem III.1]) tells us that this is equal to the total influence at p_n (defined below). Therefore the total influence dictates what the scaling factor a_n *should* be. We mention that while this heuristic for the scaling works in most natural examples, it is certainly not true in general. For example, if f_n is the Boolean function which is the AND of majority (to be defined later) on n bits and dictator, then the total influence will be of order \sqrt{n} but no scaling ($a_n \equiv 1$) is needed to obtain a nondegenerate limit for T_n .

Definition. Given a Boolean function f of n variables and a variable $i \in [n]$, we say that i is *pivotal* for f for ω if $f(\omega) \neq f(\omega^i)$ where ω^i is ω but flipped in the i th coordinate. The *influence* of the i th bit with respect to p , denoted by $\text{Inf}_i^p(f)$, is defined by

$$\text{Inf}_i^p(f) := \mathbb{P}_p(i \text{ is pivotal for } f)$$

and the *total influence* with respect to p is defined to be $\sum_{i \in [n]} \text{Inf}_i^p(f)$.

1.1. Limiting behavior for some specific Boolean functions

We now summarize some of the results that we will obtain. The paper will begin by analyzing the limiting distribution of T_n for the majority function (which will be normal), the tribes function (which will be a reverse Gumbel distribution) and certain properties associated to graphs, such as connectivity and clique containment. A connection between the tribes function and the coupon collector problem is discussed. These results, which are not difficult, are presented in Section 2.

A class of functions that offers a quite interesting analysis is the so-called iterated majority functions. For this class the analysis of the limiting distribution of T_n (both its existence and its properties) requires somewhat more work and involves dynamical systems. Given an odd integer $m \geq 3$, the iterated m -majority function is defined recursively on m^n bits as follows. One constructs an m -ary tree of height n and places 0’s and 1’s at the leaves. One takes the majority of the bits in each *family* of m leaves and thus obtains 0 and 1 values for the nodes at height $n - 1$. One then continues iteratively until the root is assigned a value. This is defined to be the output of the function.

The iterated majority function has been studied in various papers and is of interest in, among other areas, theoretical computer science. Here we simply mention one paper, by Mossel and O’Donnell [41], where these functions are explicitly studied. These authors showed that this family, as m varies, provides examples of sequences of monotone Boolean functions where the “noise sensitivity exponent” (which we do not define here) is arbitrarily close to $1/2$.

In Section 3 we identify the precise rate of decay for the tails of the limiting distributions for this class of Boolean functions. To state the result of the analysis for iterated majority, we let, for odd integers $m \geq 3$,

$$\gamma(m) := m \binom{m-1}{\frac{m-1}{2}} 2^{-(m-1)} \quad \text{and} \quad \beta(m) := \frac{\log \frac{m+1}{2}}{\log \gamma(m)}.$$

Theorem 1. Consider, for each odd integer $m \geq 3$, iterated m -ary majority on m^n bits.

- (a) Then $\gamma(m)^n (T_n - \frac{1}{2})$ converges in distribution, as n tends to infinity, to a random variable whose distribution F_m is symmetric, absolutely continuous and fully supported on \mathbb{R} . Moreover, for $m \neq m'$, F_m and $F_{m'}$ are not related by a linear change of variables.

(b) There exist constants $c_1 = c_1(m)$ and $c_2 = c_2(m)$ in $(0, \infty)$ so that for all $x \geq 1$,

$$\exp(-c_1 x^{\beta(m)}) \leq \mathbb{P}(W_m \geq x) \leq \exp(-c_2 x^{\beta(m)}),$$

where W_m has distribution F_m .

(c) $\beta(m)$ is strictly increasing, taking values in the interval $(1, 2)$ and approaches 2 as $m \rightarrow \infty$.

(d) The sequence $(F_m)_{m \geq 1}$ approaches, as $m \rightarrow \infty$, a centered Gaussian with variance $(2\pi)^{-1}$.

Remark. Note that parts (b) and (c) together state that the tails of F_m are between those of an exponential and a Gaussian. The fact that $\beta(m)$ approaches 2 is consistent with part (d).

One of the most interesting and studied sequence of Boolean functions corresponds to percolation crossings of a square. The rich structure of this particular example has inspired an extensive analysis; some parts of this recent development are presented in the book [18]. We will state below a result for this example whose proof, unlike the proofs of all other results in this paper which are proved from first principles, will be based on some recent highly nontrivial developments in percolation and in so called near-critical percolation, due to Garban, Pete and Schramm [15,16], building on work by Kesten [34], see also [43].

To even begin this, we need to introduce a number of different concepts. However, we will be very brief and refer to [53] and [17] for background and explanation of terms which are not clear. We consider percolation on the hexagonal tiling embedded into \mathbb{R}^2 . Given n , we will consider the set of hexagons contained inside of $[0, n] \times [0, n]$, denoted by B_n , and we will think of these hexagons as indexing our underlying i.i.d. random variables of which we will then have approximately n^2 . We let f_n be the indicator function of the event that there is a path of hexagons from the left side of this box to the right side all of whose values are 1. It is well known that there is a threshold at $p = 1/2$ which is the critical value for percolation on the (full) hexagonal tiling. For critical percolation on the hexagonal tiling, we let $\alpha_4(R)$ be the probability that there are four paths of alternating value from a neighbor of the origin 0 to distance R away; this event is usually called the *four-arm* event. See Figure 1 (in the appendix) for a realization of this event (where 1 is replaced by black and 0 is replaced by white). Using Schramm-Loewner evolution and conformal invariance, it was proved by Smirnov and Werner [51] that

$$\alpha_4(R) = R^{-\frac{5}{4} + o(1)} \quad \text{as } R \rightarrow \infty.$$

A little bit of thought shows that if we have a hexagon H in B_n , not too close to the boundary, which is pivotal for this crossing event, then the four-arm event to distance approximately n centered at H occurs. From here, it is possible to argue (see [17]) that the expected number of pivotal hexagons for f_n is, up to constants, $n^2 \alpha_4(n)$. This suggests what the proper scaling of T_n should be and this turns out to be correct. The following result will be proved in an appendix to this paper authored by Gábor Pete.

Theorem 2. Consider percolation crossings of an $n \times n$ -square of the hexagonal tiling.

- (a) Then $n^2 \alpha_4(n)(T_n - \frac{1}{2})$ converges in distribution, as n tends to infinity, to a random variable whose distribution F is symmetric, absolutely continuous and fully supported on \mathbb{R} .
- (b) There exist constants c_1 and c_2 in $(0, \infty)$ so that for all $x \geq 1$

$$\exp(-c_1 x^{4/3}) \leq \mathbb{P}(W \geq x) \leq \exp(-c_2 x^{4/3}),$$

where W has distribution F .

Remark. The fact that the limit in part (a) of this theorem exists follows from recent results due to Garban, Pete and Schramm [16, Theorem 1.5 and Proposition 9.6], as stated already in a previous version of this paper. However, the precise rate of decay, stated in part (b), of the tails of the limiting distribution was not known to us at the time, and only later found by the appendix author.

1.2. Limiting behavior for general Boolean functions

When one considers the question about the limiting distribution of T_n for a given sequence of Boolean functions, it is natural to ask which distributions on \mathbb{R} can arise as normalized limits of such a sequence of T_n . The next result, proved in Section 4, says that they all do. We remind the reader that a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is called transitive if it is invariant with respect to a transitive group of permutations of $[n]$.

Part (b) of the following theorem has been obtained jointly with Anders Martinsson.

Theorem 3. *Let μ denote any probability measure on \mathbb{R} .*

- (a) *For any sequence $(a_n)_{n \geq 1}$ satisfying $1 \ll a_n \ll \sqrt{n}$, there exists a sequence $(f_n)_{n \geq 1}$ of monotone functions $f_n : \{0, 1\}^n \rightarrow \{0, 1\}$ for which $a_n(T_n - \frac{1}{2})$ approaches μ in distribution.*
- (b) *For any sequence $(a_n)_{n \geq 1}$ satisfying $\log n \ll a_n \ll \sqrt{n}$, there exists a sequence $(f_n)_{n \geq 1}$ of monotone and transitive functions $f_n : \{0, 1\}^n \rightarrow \{0, 1\}$ for which $a_n(T_n - \frac{1}{2})$ approaches μ in distribution.*

Remark. By modifying the construction leading to part (b) one may obtain, for any sequence $(a_n)_{n \geq 1}$ satisfying $(\log n)^2 \ll a_n \ll n$, a sequence $(f_n)_{n \geq 1}$ of monotone graph properties, defined on n vertices and $\binom{n}{2}$ edges, for which $a_n(T_n - \frac{1}{2})$ approaches μ in distribution. Moreover, the centralizing coefficient of Theorem 3 could in greater generality be replaced by any sequence $(b_n)_{n \geq 1}$ bounded away from 0 and 1, although this may be of less interest. We further mention that Rossignol [46] has previously showed that for any sufficiently smooth sequence $(a_n)_{n \geq 1}$ satisfying $\log n \leq a_n \leq \sqrt{n}$ there exists an increasing sequence $(N(n))_{n \geq 1}$ and monotone and transitive Boolean functions $(f_{N(n)})_{n \geq 1}$ with a threshold at $1/2$ of width $1/a_n$.

There are some conditions that the scaling coefficients $(a_n)_{n \geq 1}$ have to meet in order to obtain a non-degenerate limit. Further restrictions apply in order not to impose properties on the limiting distribution. These facts are described in the following proposition. While some of these facts are well known to many we present them here for convenience to the reader, and provide a proof in Section 4. Together they show that Theorem 3 is in fact sharp.

Proposition 4. *Assume that $a_n(T_n - b_n)$ converges, as n tends to infinity, to some non-degenerate probability measure μ , and that $(b_n)_{n \geq 1}$ is bounded away from 0 and 1. Then:*

- (a) *The sequence $(a_n/\sqrt{n})_{n \geq 1}$ is bounded from above.*
- (b) *If the functions $(f_n)_{n \geq 1}$ are transitive, then $(a_n/\log n)_{n \geq 1}$ is bounded away from 0.*
- (c) *If $(a_n)_{n \geq 1}$ is bounded from above, then μ is necessarily fully supported on a (finite) interval.*
- (d) *If the functions $(f_n)_{n \geq 1}$ are transitive and $(a_n/\log n)_{n \geq 1}$ is bounded from above, then μ is necessarily fully supported on a (possibly infinite) interval.*
- (e) *If $(a_n/\sqrt{n})_{n \geq 1}$ is bounded away from 0, then μ is necessarily absolutely continuous.*

Remark. Part (e) of the above proposition was pointed out to us by Anders Martinsson. We also mention that the statement in the previous remark regarding graph properties is also essentially best possible, due to the work of Bourgain and Kalai [7].

Interestingly, there are sequences of nondegenerate random variables $(X_n)_{n \geq 1}$ which are not *renormalizable* in the sense that for no subsequence $(X_{n_k})_{k \geq 1}$ are there normalizing constants $(a_k)_{k \geq 1}$ and $(b_k)_{k \geq 1}$ with the a_k 's nonnegative so that $a_k(X_{n_k} - b_k)$ converges, as $k \rightarrow \infty$, to a nondegenerate limiting distribution. A typical example of such a sequence is given by $X_n = e^{nZ}$ where Z is a standard normal random variable. The vague idea is that when we try to scale down to keep mass from going to infinity, then the result will be that all the mass is accumulating at 0. Another example, which we will exploit, is when X_n is uniformly distributed on $\{\pm 2^k : k = 1, 2, \dots, n\}$.

The following proposition shows that one cannot necessarily extract a subsequence of $(T_n)_{n \geq 1}$ which after normalization converges to a nondegenerate limit.

Proposition 5. *There exists a (nondegenerate) sequence $(f_n)_{n \geq 1}$ of monotone (and transitive) Boolean functions $f_n : \{0, 1\}^n \rightarrow \{0, 1\}$ so that no subsequence of $(T_n)_{n \geq 1}$ can be renormalized to have a nondegenerate limiting distribution.*

Remark. The proof of this result, provided in Section 4, will be based on Theorem 3. We may therefore assume that the functions in the obtained sequence are transitive if we so wish.

We end this introductory section with an open question. As the reader may notice, all examples we have worked out yield a limiting distribution with exponentially or super-exponentially decaying tails. Although our Theorem 3 certainly shows that there are (sequences of) monotone Boolean functions giving rise to limiting distributions with heavier tails, the examples we construct are not very natural. Are there any “natural” examples whose limiting distributions present sub-exponential tails?

2. Some elementary examples

2.1. Majority and the standard normal

Majority is an example providing nontrivial, although classical, scaling behavior. The majority function on n bits is defined to output the value 1 if there are at least $n/2$ bits with the value 1. More generally, we consider biased majority, which is the function with output 1 if and only if there are at least pn bits valued 1, where $p \in (0, 1)$ is a fixed parameter. The correct scaling factor will be of order \sqrt{n} and the limit will be Gaussian, as stated in the following proposition.

Proposition 6. *For every $p \in (0, 1)$ we have for the p -biased majority function on n bits that $\sqrt{\frac{n}{p(1-p)}}(T_n - p)$ converges in distribution to a standard normal.*

Note that the multiplicative scaling is of order \sqrt{n} and coincides with the order of the total influence at the relevant parameter p ; hence it is consistent with the heuristic described above.

Proof. Let $x \in \mathbb{R}$, $a_n = \sqrt{n/[p(1-p)]}$ and $p_n = p + x/a_n$. For large n we have $p_n \in [0, 1]$, and

$$\mathbb{P}(a_n(T_n - p) \leq x) = \mathbb{P}(T_n \leq p_n) = \mathbb{P}_{p_n} \left(\sum_{i=1}^n \omega_i \geq np \right).$$

We of course have a sum of n Bernoulli variables with success probabilities p_n .

A consequence of the Lindeberg-Feller central limit theorem (see e.g. [9, Theorem 3.4.5]) is that if $\{X_{i,n} : 1 \leq i \leq n, n \geq 1\}$ is a family of bounded random variables, such that for each n , $\{X_{i,n} : 1 \leq i \leq n\}$ are i.i.d. with zero mean and variance that tends to 1 as n increases, then $\sum_{i=1}^n X_{i,n}/\sqrt{n}$ converges in distribution to a standard normal.

Since $\text{Var}_{p_n}(\omega_i) = p_n(1-p_n)$, which tends to $p(1-p)$, and $(np - np_n)/\sqrt{np(1-p)} = -x$, the above consequence of the Lindeberg-Feller theorem implies that, as $n \rightarrow \infty$,

$$\mathbb{P}_{p_n} \left(\sum_{i=1}^n \omega_i \geq np \right) = \mathbb{P}_{p_n} \left(\sum_{i=1}^n \frac{\omega_i - p_n}{\sqrt{np(1-p)}} \geq \frac{np - np_n}{\sqrt{np(1-p)}} \right) \rightarrow \Phi(x),$$

the distribution function of a standard normal distribution. □

2.2. Tribes, Gumbel and coupon collectors

The tribes function on n bits is defined as follows. Given ℓ_n , partition $[n]$ into $\lfloor n/\ell_n \rfloor$ sets (‘tribes’) of length ℓ_n (plus some residual bits). Then $f_n(\omega) = 1$ if and only if ω is all 1’s for at least one tribe. The *correct* choice for ℓ_n , in order for the distribution of to be nondegenerate for the uniform measure, is of order $\log_2 n - \log_2 \log_2 n$.

Proposition 7. *Consider tribes with $\ell_n = \lfloor \log_2 n - \log_2 \log_2 n \rfloor$, set $\alpha_n = (\log_2 n - \log_2 \log_2 n)/\ell_n$. Then for all $x \in \mathbb{R}$ we have*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(2(\log_2 n) \left(T_n - \left(\frac{1}{2} \right)^{\alpha_n} \right) \leq x \right) \rightarrow 1 - \exp(-e^x).$$

Note that the multiplicative scaling is of order $\log_2 n$, which can be checked to be the order of the total influence at the relevant parameter $1/2$; again, this is consistent with the heuristic described in the introduction. Note also that the upper tail of this limiting distribution decays super-exponentially, whereas the lower tail just decays exponentially.

Proof. Note that the $\lfloor n/\ell_n \rfloor$ tribes of length ℓ_n corresponds to the 1-witnesses for the tribes function. Let X_n denote the number of tribes (1-witnesses) for which ω is all 1. For $\omega \sim \mathbb{P}_p$, we see that X_n is binomially distributed with parameters $\lfloor n/\ell_n \rfloor$ and p^{ℓ_n} . Clearly, $\lfloor n/\ell_n \rfloor \rightarrow \infty$ as $n \rightarrow \infty$. Also, for all $x \in \mathbb{R}$ we have

$$\lfloor n/\ell_n \rfloor \left(\left(\frac{1}{2} \right)^{\alpha_n} + \frac{x}{2 \log_2 n} \right)^{\ell_n} = \lfloor n/\ell_n \rfloor \left(\frac{1}{2} \right)^{\alpha_n \ell_n} \left(1 + \frac{x}{2^{1-\alpha_n} \log_2 n} \right)^{\ell_n} \rightarrow e^x$$

as $n \rightarrow \infty$. Given $x \in \mathbb{R}$ and letting $p_n = \left(\frac{1}{2} \right)^{\alpha_n} + \frac{x}{2 \log_2 n}$, we therefore have, by the Poisson convergence theorem (see e.g. [9, Theorem 3.6.1]), that for $\omega \sim \mathbb{P}_{p_n}$, $X_n(\omega)$ converges in law to a Poisson distribution with parameter e^x . Since for each n ,

$$\mathbb{P} \left(T_n \leq \left(\frac{1}{2} \right)^{\alpha_n} + \frac{x}{2 \log_2 n} \right) = \mathbb{P}_{p_n} (X_n \geq 1),$$

we thus conclude that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(T_n \leq \left(\frac{1}{2} \right)^{\alpha_n} + \frac{x}{2 \log_2 n} \right) = 1 - \exp(-e^x),$$

as we needed to show. □

Remark. The unfortunate term α_n arises due to the fact that $\log_2 n - \log_2 \log_2 n$ is not an integer. A related fact is that if $p = 1/2$, then the number of tribes which are identically 1 has all Poisson distributions with parameter in $[1, 2]$ as subsequential limits. If we were to restrict ourselves to n 's of the form 2^{2^k} , then $\log_2 n - \log_2 \log_2 n$ would be an integer and we would have the simpler form that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(2(\log_2 n) \left(T_n - \frac{1}{2} \right) \leq x \right) \rightarrow 1 - \exp(-e^x)$$

along this thin subsequence of n .

The reader might recognize the limiting distribution obtained in Proposition 7. In general if X has distribution $F(x)$, then $-X$ has distribution $1 - F(-x)$. If Y is distributed according to the above limiting distribution, then $-Y$ has distribution $\exp(-e^{-x})$ which is known as the standard Gumbel distribution. This distribution often arises in extreme value theory and in particular is the limiting distribution after proper normalization of

- (a) the maximum of n independent unit exponential random variables (where one subtracts $\log n$ but uses no scaling factor to normalize); and
- (b) the number of picks needed to collect n coupons when each pick is uniform (where one subtracts $n \log n$ and divides by n to normalize).

Heuristically, the reason that one gets the same limiting distribution in these two models is that in the latter case, we have the maximum of n weakly dependent geometric random variables with parameters $1/n$. When dividing by n (which explains the difference of a factor of n in the two normalizations), the geometric random variables become unit exponentials in the limit.

While we will not give an alternative proof of Proposition 7 based on these ideas, we want to *explain* why we obtained the limiting distribution there that we did.

Given a Boolean function f , define its reversal \hat{f} by $\hat{f}(\omega) = 1 - f(1 - \omega)$ and observe that \hat{f} is also a monotone Boolean function. One immediately checks that $T(\hat{f})$ and $1 - T(f)$ have the same distribution. If f_n is our tribes

function, then this distributional relationship and Proposition 7 easily yields that

$$2(\log_2 n) \left(T(\hat{f}_n) - 1 + \left(\frac{1}{2}\right)^{\alpha_n} \right) \tag{3}$$

converges to the standard Gumbel distribution. We now give a heuristic for this. Clearly, \hat{f}_n is the function which is 0 if and only if there is a tribe which is all 0's. (The tribes are 0-witnesses for \hat{f}_n .) One easily checks that $T(\hat{f}_n)$ is the smallest p such that each tribe has a 1 in it with respect to η_p ; compare with (2). The distribution of the time at which a given tribe gets its first 1 is equal to the distribution of the minimum of ℓ_n uniform random variables. The minimum of k uniform random variables after multiplying by k converges to a unit exponential. Therefore, since different tribes are disjoint (and hence their corresponding uniform random variables are independent) and have size ℓ_n , it follows that $\ell_n T(\hat{f}_n)$ is approximately the maximum of $\lfloor n/\ell_n \rfloor$ unit exponential random variables. Therefore one *should* have that

$$\ell_n T(\hat{f}_n) - \log(\lfloor n/\ell_n \rfloor) = \ell_n \left(T(\hat{f}_n) - \frac{\log(\lfloor n/\ell_n \rfloor)}{\ell_n} \right)$$

converges to the Gumbel distribution. This is certainly close to (3) and heuristically explains the *reverse* Gumbel distributional limit.

Remark. The so-called circular tribes function is a more symmetric version of tribes and perhaps more natural. It is defined as follows. We place the n bits in a circle and define $f_n(\omega)$ to be 1 if ω contains an interval of 1's of length $\lfloor \log_2 n \rfloor$. One can prove in a similar manner that the corresponding sequence T_n also has the reverse Gumbel distribution as a limit. The situation is however slightly different than for tribes since the number of such intervals containing all 1's is no longer Poisson but rather compound Poisson, where the summands are mean 2 geometric random variables.

2.3. Random graph properties

In this subsection we cover a few monotone functions related to random graphs. We remind the reader that a random graph on n vertices is obtained by declaring each of the possible $\binom{n}{2}$ edges open with probability $p \in (0, 1)$. Equivalently, this amounts to determining an element $\omega \in \{0, 1\}^{\binom{n}{2}}$ according to \mathbb{P}_p . We first discuss two functions whose critical values occur near 0. The proof of the following proposition is very straightforward (when using well known results) and hence we only sketch the proof.

Proposition 8.

(a) *Let f_n be the function corresponding to containing a triangle in a graph with n vertices. Then, for all $x \geq 0$, we have that*

$$\lim_{n \rightarrow \infty} \mathbb{P}(nT_n \leq x) = 1 - \exp(-x^3/6).$$

(b) *Let f_n be the function corresponding to a graph with n vertices being connected. Then, for all $x \in \mathbb{R}$, we have that*

$$\lim_{n \rightarrow \infty} \mathbb{P}(nT_n - \log n \leq x) = \exp(-e^{-x}).$$

The multiplicative scaling n can in both cases be checked to be the order of the total influence at the relevant parameter.

Proof. (a) It is well known (see e.g. [3, Theorem 4.1]) that if $p = x/n$, then the number of triangles contained in the random graph converges to a Poisson distribution with parameter $x^3/6$. The result follows immediately using (1).

(b) It is well known (see e.g. [3, Theorem 7.3]) that for any $x \in \mathbb{R}$, if $p = (\log n + x)/n$, the probability that the random graph is connected approaches $\exp(-e^{-x})$. The result follows. □

Remark. We see that in part (a) of the above proposition the threshold is coarse and the support of the limiting distribution bounded to the left, while in part (b) the threshold is sharp and the support unbounded to the left. This is an instance of the general phenomenon that only sharp thresholds may give rise to limiting distributions supported on the whole line. More precisely, assume that $(b_n)_{n \geq 1}$ is bounded away from 1 and that $a_n(T_n - b_n)$ converges, as n tends to infinity, to some probability measure μ . Then a coarse threshold is characterized by the sequence $(a_n b_n)_{n \geq 1}$ being bounded above, and if c is an upper bound on this sequence, then the support of μ is contained in $[-c, \infty)$.

A *clique* is a maximal complete subgraph of a graph. At a given parameter $p \in (0, 1)$, the expected number of complete subgraphs of size ℓ of a random graph on n vertices falls abruptly from being very large to being very small, as ℓ increases. As a consequence, the maximal clique size of a random graph is highly concentrated, with high probability equal to either of two consecutive values $\ell_n - 1$ or ℓ_n , where $\ell_n = \ell_n(p)$. Using Stirling's approximation one sees that this sequence must satisfy $\ell_n \sim 2 \log_{1/p} n$. This is well known; see e.g. [3, Chapter 4].

We will be interested in the function encoding the existence of a clique of size ℓ_n . For most values of n the maximal sized clique consists of ℓ_n vertices with probability close to 1. However, along certain subsequences this probability remains bounded away from 1. Instead of restricting to subsequences we may allow p to vary, similar to the case of tribes. We simply state this result without proof since the argument follows more or less the argument for tribes. One obtains the result by proving Poisson approximation for the number of complete graphs of a given size. While this is more involved than for tribes, it is proved in [3, Theorems 11.7 and 11.9].

Proposition 9. *Let $p \in (0, 1)$ and $\ell_n = \ell_n(p)$ be the above mentioned sequence. Let p_1, p_2, \dots be any sequence bounded away from 0 and 1 such that the limit*

$$\lambda := \lim_{n \rightarrow \infty} \binom{n}{\ell_n} p_n^{\binom{\ell_n}{2}} \quad \text{exists in } (0, \infty).$$

Then, for the Boolean function encoding the existence of a complete graph of size ℓ_n , we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\ell_n^2}{2p_n} (T_n - p_n) \leq x \right) = 1 - \exp(-\lambda e^x) \quad \text{for } x \in \mathbb{R}.$$

3. Iterated majority

In this section, we will analyze iterated majority and prove Theorem 1. In order to understand the asymptotic behavior of iterated majority, one is led to study its recursive structure. The limiting distribution will be described through the iterates of some function $g : [0, 1] \rightarrow [0, 1]$, and the appropriate scaling will be determined by the derivative of g at $\frac{1}{2}$.

We begin by describing what the limiting distribution F_m will be. Define $g : [0, 1] \rightarrow [0, 1]$ as the probability at parameter $x \in [0, 1]$ that the majority on m bits equals 1. Formally, g is given by

$$g(x) = \sum_{k=(m+1)/2}^m \binom{m}{k} x^k (1-x)^{m-k}.$$

Observe that $\gamma(m)$, which will be our scaling coefficient satisfies

$$\gamma(m) := \frac{m}{m-1} \gamma(m-2). \tag{4}$$

It is clear that $\gamma(m)$ is increasing in m , and Stirling's approximation says that $\gamma(m) \sim \sqrt{2m/\pi}$ as m tends to infinity. It turns out that the total influence for f_n , i.e. iterated majority on m^n bits (where m is implicit), is γ^n and we will below see that γ coincides with the derivative of g at $\frac{1}{2}$. The recursive structure of $\gamma(m)$ stated in (4) easily yields that $\gamma(m) < \frac{m}{2}$, implying in turn that $\beta(m) > 1$ for all $m \geq 3$. Also, using $\gamma(m) \sim \sqrt{2m/\pi}$, we find that $\beta(m) \rightarrow 2$ as $m \rightarrow \infty$.

It turns out to be convenient to consider the translate

$$h(x) = g\left(\frac{1}{2} + x\right) - \frac{1}{2}$$

of g . The scaling limit of iterated m -majority will be described in terms of the limit as $n \rightarrow \infty$ of $h^{(n)}(\alpha\gamma^{-n})$, where $\alpha \in \mathbb{R}$ and $h^{(n)}$ denotes the composition of h with itself n times.

We will break up the proof of the four parts of Theorem 1 into subsections.

3.1. Proof of part (a)

We begin with the following proposition which will be central for our analysis.

Proposition 10. *For every odd integer $m \geq 3$ and $\alpha \in \mathbb{R}$, the limit $L(\alpha) := \lim_{n \rightarrow \infty} h^{(n)}(\alpha\gamma^{-n})$ exists and the resulting function $L : \mathbb{R} \rightarrow (-\frac{1}{2}, \frac{1}{2})$ is odd, onto, 1-Lipschitz continuous, strictly increasing and continuously differentiable.*

We first need the following lemma.

Lemma 11. *The function $h : [-\frac{1}{2}, \frac{1}{2}] \rightarrow [-\frac{1}{2}, \frac{1}{2}]$ is odd, onto, strictly increasing, strictly convex on $[-\frac{1}{2}, 0]$ and strictly concave on $[0, \frac{1}{2}]$. In particular $h'(x) \leq h'(0) = \gamma(m)$ for $x \in [-\frac{1}{2}, \frac{1}{2}]$.*

Proof. It suffices to demonstrate the corresponding characteristics for g . From the interpretation of g as a probability, it is clear that g is strictly increasing, maps $0, \frac{1}{2}$ and 1 to themselves, and that

$$g(x) = \mathbb{P}(\text{Bin}(m, x) \geq m/2) = 1 - \mathbb{P}(\text{Bin}(m, x) < m/2) = 1 - g(1 - x).$$

Thus h is odd, strictly increasing, has fixed points at $-\frac{1}{2}, 0$ and $\frac{1}{2}$, and is therefore also onto.

We know that g is differentiable and we aim to determine its derivative. Note that

$$g'(1) = g'(0) = \lim_{x \rightarrow 0} \sum_{k \geq m/2} \binom{m}{k} x^{k-1} (1-x)^{m-k} = 0.$$

Next, pick $\delta > 0$ and let $\xi_1, \xi_2, \dots, \xi_m$ be independent and $[0, 1]$ -uniformly distributed. Using the monotone coupling we find that

$$g(x + \delta(1-x)) - g(x) = \mathbb{P}(\#\{\xi_i \leq x\} < m/2, \#\{\xi_i \leq x + \delta(1-x)\} \geq m/2).$$

Conditioning on the number of ξ_i 's whose value is at most x we arrive at

$$\sum_{k \leq m/2} \binom{m}{k} x^k (1-x)^{m-k} \mathbb{P}(\#\{\xi_i \leq x + \delta(1-x)\} \geq m/2 | \#\{\xi_i \leq x\} = k).$$

The above conditional probability coincides with the probability that a binomial random variable with parameters $m - k$ and δ is at least $m/2 - k$, and is thus independent of x . In addition,

$$\delta^{-1} \mathbb{P}(\text{Bin}(m - k, \delta) \geq m/2 - k) = \sum_{\ell \geq m/2 - k} \binom{m - k}{\ell} \delta^{\ell-1} (1 - \delta)^{m-k-\ell},$$

and sending δ to 0 leaves us with $m - k = (m + 1)/2$ in case $k = (m - 1)/2$, and 0 for all smaller values of k . In conclusion, for $x \in (0, 1)$,

$$g'(x) = \lim_{\delta \rightarrow 0} \frac{g(x + \delta(1-x)) - g(x)}{\delta(1-x)} = \frac{m+1}{2} \binom{m}{\frac{m-1}{2}} [x(1-x)]^{\frac{m-1}{2}}. \tag{5}$$

Differentiating once more gives

$$g''(x) = \frac{m+1}{2} \frac{m-1}{2} \binom{m}{\frac{m-1}{2}} [x(1-x)]^{\frac{m-3}{2}} (1-2x).$$

In conclusion, the derivative of g is strictly positive on $(0, 1)$, and the second derivative is strictly positive on $(0, \frac{1}{2})$ and strictly negative on $(\frac{1}{2}, 1)$. So h possesses the claimed properties and h' reaches its maximum at the origin, which is easily seen to equal $\gamma(m)$. \square

The proof of Proposition 10 will make repeated use of the properties of h displayed in Lemma 11. For instance, we note that h cannot have any fixed points other than $-\frac{1}{2}$, 0 and $\frac{1}{2}$.

Proof of Proposition 10. Since $h(0) = 0$ we also have $L(0) = 0$, and since h is odd the limit $L(\alpha)$, if it exists, has to be odd as well. In particular, it will be sufficient to consider $\alpha \geq 0$ for the rest of this proof.

Existence. Given $\alpha \geq 0$, choose n_0 such that $\alpha\gamma^{-n} \leq \frac{1}{2}$ for all $n \geq n_0$. Note that we may obtain $h^{(n)}(\alpha\gamma^{-n})$ from $\alpha\gamma^{-(n+1)}$ by first multiplying by γ , and then applying h n times. $h^{(n+1)}(\alpha\gamma^{-(n+1)})$ is similarly obtained from $\alpha\gamma^{-(n+1)}$ by first applying h once, and then another n times. Lemma 11 shows that the derivative of h is bounded by γ . Hence $\gamma x \geq h(x)$ for all $x \in [0, 1/2]$, and it follows that $h^{(n)}(\alpha\gamma^{-n})$ is decreasing in n for $n \geq n_0$. Since the sequence is bounded below by 0 , the limit $L(\alpha)$ necessarily exists for all $\alpha \geq 0$.

1-Lipschitz Continuity. Using again that $|h'| \leq \gamma$, together with iterated use of the mean value theorem, we find for $\alpha, \alpha' \in \mathbb{R}$ that

$$|L(\alpha) - L(\alpha')| = \lim_{n \rightarrow \infty} |h^{(n)}(\alpha\gamma^{-n}) - h^{(n)}(\alpha'\gamma^{-n})| \leq \liminf_{n \rightarrow \infty} \gamma^n |\alpha\gamma^{-n} - \alpha'\gamma^{-n}| = |\alpha - \alpha'|,$$

where we also have used that $\alpha\gamma^{-n}$ and $\alpha'\gamma^{-n}$ are contained in $[-\frac{1}{2}, \frac{1}{2}]$ for large n .

An observation that will be important for the rest of this proof is that, by continuity of h , for all $\alpha \in \mathbb{R}$

$$h(L(\alpha)) = \lim_{n \rightarrow \infty} h^{(n+1)}(\alpha\gamma^{-n}) = L(\alpha\gamma). \quad (6)$$

Iterating this yields that

$$L(\alpha) = h(L(\alpha\gamma^{-1})) = h^{(n)}(L(\alpha\gamma^{-n})). \quad (7)$$

Strict Monotonicity. Note that weak monotonicity of course follows from h being increasing. We will next aim to show that for $\alpha \geq \alpha' \geq 0$ sufficiently small, we have

$$L(\alpha) - L(\alpha') \geq (\alpha - \alpha') \prod_{k=1}^{\infty} \left(1 - \alpha \left(\frac{3}{4}\right)^k\right). \quad (8)$$

Apart from showing that L is strictly increasing in a neighborhood around the origin, (8) will be an important step in the proof of differentiability of L . Note that strict monotonicity of L would follow for all $\alpha \in \mathbb{R}$ by (7) and (8), since h is strictly increasing.

We now deduce (8). Using concavity of h on $[0, \frac{1}{2}]$, we observe that $\frac{4}{3} \leq \gamma - \frac{1}{6} \leq h'(x) \leq \gamma$ on some interval $[0, c]$, where $c = c(m) > 0$. So, by the mean value theorem we conclude that

$$\frac{4}{3}x \leq h(x) \leq \gamma x \quad \text{on } [0, c].$$

Consequently, for all $\alpha \in [0, c]$ and $1 \leq k \leq n$, we have $h^{(k)}(\alpha\gamma^{-n}) \leq \alpha$, and therefore

$$h^{(k)}(\alpha\gamma^{-n}) \leq \frac{3}{4} h^{(k+1)}(\alpha\gamma^{-n}) \leq \left(\frac{3}{4}\right)^{n-k} h^{(n)}(\alpha\gamma^{-n}) \leq \alpha \left(\frac{3}{4}\right)^{n-k}.$$

Now, for any $\alpha' \leq \alpha$ in $[0, c]$ and given n , we obtain, from iterated use of the mean value theorem, the existence of constants $\{s_k^n\}_{1 \leq k \leq n}$ with $s_k^n \in [h^{(k-1)}(\alpha'\gamma^{-n}), h^{(k-1)}(\alpha\gamma^{-n})]$ and such that

$$h^{(n)}(\alpha\gamma^{-n}) - h^{(n)}(\alpha'\gamma^{-n}) = (\alpha\gamma^{-n} - \alpha'\gamma^{-n}) \prod_{k=1}^n h'(s_k^n). \tag{9}$$

Since h' is decreasing on $[0, \frac{1}{2}]$ and $h''(0) = 0$, we have that $h'(x)$ is bounded below by $\gamma(1 - x)$ on some, possibly smaller, interval $[0, c']$. As a consequence we obtain the lower bound on (9),

$$\begin{aligned} & (\alpha\gamma^{-n} - \alpha'\gamma^{-n}) \prod_{k=1}^n \gamma(1 - h^{(k-1)}(\alpha\gamma^{-n})) \\ & \geq (\alpha - \alpha') \prod_{k=1}^n \left(1 - \alpha \left(\frac{3}{4}\right)^{n-k+1}\right) \\ & = (\alpha - \alpha') \prod_{k=1}^n \left(1 - \alpha \left(\frac{3}{4}\right)^k\right) \geq (\alpha - \alpha') \prod_{k=1}^{\infty} \left(1 - \alpha \left(\frac{3}{4}\right)^k\right). \end{aligned} \tag{10}$$

Combining (9) and (10) and letting $n \rightarrow \infty$ yields (8) for every $\alpha' \leq \alpha$ in $[0, c']$.

Continuous Differentiability. Using (7) we have for any $\alpha \geq 0$ and $\delta \in \mathbb{R}$ that

$$\begin{aligned} L(\alpha + \delta) - L(\alpha) &= h^{(n)}(L((\alpha + \delta)\gamma^{-n})) - h^{(n)}(L(\alpha\gamma^{-n})) \\ &= [L((\alpha + \delta)\gamma^{-n}) - L(\alpha\gamma^{-n})] \prod_{k=1}^n h'(h^{(k-1)}(L(\alpha_k\gamma^{-n}))) \end{aligned}$$

and where we in the last step have used the mean value theorem iteratively; the α_k 's are bounded between α and $\alpha + \delta$. By continuity and monotonicity of h and L , these α_k 's exist. Using $h^{(k-1)}(L(\alpha)) = L(\alpha\gamma^{k-1})$, and reindexing the terms of the product, we arrive at

$$\frac{L(\alpha + \delta) - L(\alpha)}{\delta} = \frac{L((\alpha + \delta)\gamma^{-n}) - L(\alpha\gamma^{-n})}{\delta\gamma^{-n}} \prod_{k=1}^n \gamma^{-1} h'(L(\alpha_{n-k+1}\gamma^{-k})). \tag{11}$$

We now want to take limits. First, since L is 1-Lipschitz continuous,

$$\limsup_{\delta \rightarrow 0} \frac{L(\alpha + \delta) - L(\alpha)}{\delta} \leq \prod_{k=1}^n \gamma^{-1} h'(L(\alpha\gamma^{-k})),$$

which is decreasing in n . Second, we note that the infinite product in (8) tends to 1 as $\alpha \rightarrow 0$. Applying this to the first term in (11), we conclude that for every $\varepsilon > 0$, if n is sufficiently large, then

$$\liminf_{\delta \rightarrow 0} \frac{L(\alpha + \delta) - L(\alpha)}{\delta} \geq (1 - \varepsilon) \prod_{k=1}^n \gamma^{-1} h'(L(\alpha\gamma^{-k})).$$

Sending n to infinity, and then ε to zero, we conclude that the inferior and superior limits coincide and that

$$L'(\alpha) = \lim_{\delta \rightarrow 0} \frac{L(\alpha + \delta) - L(\alpha)}{\delta} = \prod_{k=1}^{\infty} \gamma^{-1} h'(L(\alpha\gamma^{-k})). \tag{12}$$

Since $h' \leq \gamma$ the limit is finite, and since $L(\alpha\gamma^{-k}) \leq \alpha\gamma^{-k}$ and $h'(x) \geq \gamma(1 - x)$ for small $x \geq 0$, the limit is strictly positive for all $\alpha \in \mathbb{R}$. This, again, shows that L is strictly monotone on \mathbb{R} .

We need to show that L' is continuous, and note, based on (12), that L' is decreasing on $[0, \infty)$ since L is increasing. Since also $L' > 0$ on \mathbb{R} it follows that

$$\prod_{k=\ell}^{\infty} \gamma^{-1} h'(L(\alpha \gamma^{-k})) \rightarrow 1 \quad \text{as } \ell \rightarrow \infty$$

uniformly on compact sets. Thus, for every $\varepsilon > 0$

$$\left| \lim_{x \rightarrow \alpha} L'(x) - \prod_{k=1}^{\ell} \gamma^{-1} h'(L(\alpha \gamma^{-k})) \right| < \varepsilon$$

for large enough ℓ , showing that $\lim_{x \rightarrow \alpha} L'(x) = L'(\alpha)$.

Surjectivity. Since $L(0) = 0$ and L is continuous, it remains to show that $L(\alpha) \rightarrow \frac{1}{2}$ as $\alpha \rightarrow \infty$. For any $k \in \mathbb{N}$ we have from (7) that $L(\alpha \gamma^k) = h^{(k)}(L(\alpha))$. Since $L(\alpha) > 0$ for $\alpha > 0$, the properties of h imply that $h^{(k)}(L(\alpha)) \rightarrow \frac{1}{2}$ as $k \rightarrow \infty$. Together with the proven continuity of L and its weak monotonicity, this shows that L maps $[0, \infty)$ onto $[0, \frac{1}{2})$. \square

We now use the above to analyze the asymptotics of T_n for iterated m -majority on m^n bits to prove part (a) of Theorem 1. With $x \in \mathbb{R}$ fixed, the goal will be to relate, for large n , the probability $\mathbb{P}(T_n \leq p_n)$, where $p_n = \frac{1}{2} + x\gamma^{-n}$, with the function $L : \mathbb{R} \rightarrow [-\frac{1}{2}, \frac{1}{2}]$ defined in Proposition 10.

Given $p \in (0, 1)$ and $n \in \mathbb{N}$, let

$$q_n(p) := \mathbb{P}_p(f_n(\omega) = 1).$$

Using the iterated majority structure, it is easy to express $q_n(p)$ in terms of $q_{n-1}(p)$. Specifically, $q_n(p)$ is the probability that a binomial random variable with parameters m and $q_{n-1}(p)$ is at least $m/2$. That is,

$$q_n(p) = \mathbb{P}(\text{Bin}(m, q_{n-1}(p)) \geq m/2) = g(q_{n-1}(p)),$$

where g is defined as above. Of course, the base case $q_0(p)$ equals the probability of the majority function on one bit being 1, which equals p . Iterating the above procedure we get $q_n(p) = g^{(n)}(p)$. Replacing p by $p_n = \frac{1}{2} + x\gamma^{-n}$ leaves

$$\mathbb{P}(T_n \leq p_n) = \mathbb{P}_{p_n}(f_n(\omega) = 1) = q_n(p_n) = g^{(n)}\left(\frac{1}{2} + x\gamma^{-n}\right) = \frac{1}{2} + h^{(n)}(x\gamma^{-n}),$$

which, according to Proposition 10, converges as $n \rightarrow \infty$ to $\frac{1}{2} + L(x)$.

That the limiting distribution is absolutely continuous and fully supported on \mathbb{R} is a consequence of the properties of L established in Proposition 10. The final sentence of part (a) follows from parts (b) and (c).

Remark. Observe that the first equality in (7) has a very nice interpretation. It says that the two homeomorphisms $x \mapsto \gamma x$ from \mathbb{R} to itself and $x \mapsto h(x)$ from $[-1/2, 1/2]$ to itself are conjugate and that L provides a conjugation between these homeomorphisms (showing that they are conjugate). We know that L is also a homeomorphism. Since the two conjugate mappings are analytic, one might guess that one can show that L has good properties (such as analyticity or being continuously differentiable) by virtue of the fact it is a conjugacy between these systems. This is unfortunately not true. One can construct self-conjugacies f of $x \mapsto \gamma x$ which, while being homeomorphisms, are not continuously differentiable. Being a self-conjugacy amounts to saying that $f(\gamma x) = \gamma f(x)$ and so in particular we are saying that $f(\gamma x) = \gamma f(x)$ does not imply that f is linear even for homeomorphisms. Since $L \circ f$ would also be a conjugacy between the two systems, one cannot conclude good properties of L using only the fact that it is a conjugacy.

3.2. *Proof of part (b)*

We first need the following lemma.

Lemma 12. *For every odd integer $m \geq 3$ there exists $\varepsilon_0 = \varepsilon_0(m) > 0$ such that for all $\ell \geq 1$ and $\varepsilon \in (0, \varepsilon_0)$,*

$$\varepsilon^{\binom{m+1}{2}^\ell} \leq \left| h^{(\ell)}\left(\frac{1}{2} - \varepsilon\right) - \frac{1}{2} \right| \leq (9\varepsilon)^{\binom{m+1}{2}^\ell}.$$

Proof. First note that

$$\frac{1}{2} - h\left(\frac{1}{2} - \varepsilon\right) = 1 - g(1 - \varepsilon) = g(\varepsilon),$$

which after iteration leaves us with $\frac{1}{2} - h^{(\ell)}\left(\frac{1}{2} - \varepsilon\right) = g^{(\ell)}(\varepsilon)$. In addition, each term in g with non-zero coefficient has degree at least $(m + 1)/2$. Thus,

$$g(x) = \binom{m}{\frac{m+1}{2}} x^{\frac{m+1}{2}} (1 + r(x))$$

for some polynomial $r(x) \rightarrow 0$ as $x \rightarrow 0$. Since

$$\binom{m}{\frac{m+1}{2}} = \binom{m}{\frac{m-1}{2}} \leq \binom{2em}{m-1}^{\frac{m-1}{2}} \leq 9^{\frac{m-1}{2}},$$

we see that $g(\varepsilon)$ lies in $[\varepsilon^{\frac{m+1}{2}}, (9\varepsilon)^{\frac{m+1}{2}}/9]$ for all sufficiently small ε . Using the fact that g is increasing and $g(x) \leq x$ on $(0, \frac{1}{2})$, it then follows by two inductions that $\varepsilon^{\binom{m+1}{2}^\ell}$ and $(9\varepsilon)^{\binom{m+1}{2}^\ell}/9$ are lower respectively upper bounds for $g^{(\ell)}(\varepsilon)$. □

Now, fix $m \geq 3$, and let $\varepsilon_0 = \varepsilon_0(m) > 0$ be given as in Lemma 12. We first show the second inequality. Since L approaches $\frac{1}{2}$ continuously, we can choose $a_0 > 0$ so that $L(a_0) = \frac{1}{2} - \frac{\varepsilon_0}{9}$. Given $x \geq 1$, let $n_x := \lfloor \log_\gamma \frac{x}{a_0} \rfloor$. We first restrict to x 's which are sufficiently large so that $n_x \geq 1$. This immediately yields

$$a_0 \leq x\gamma^{-n_x}.$$

Using (7) and monotonicity of L and h , we have

$$L(x) = h^{(n_x)}(L(x\gamma^{-n_x})) \geq h^{(n_x)}(L(a_0)).$$

By Lemma 12 and the definition of a_0 , we have that

$$\mathbb{P}(W_m \geq x) = \frac{1}{2} - L(x) \leq \frac{1}{2} - h^{(n_x)}(L(a_0)) = \frac{1}{2} - h^{(n_x)}\left(\frac{1}{2} - \frac{\varepsilon_0}{9}\right) \leq \varepsilon_0^{\binom{m+1}{2}^{n_x}}.$$

An easy computation shows that, for all x for which $n_x \geq 1$

$$\left(\frac{m+1}{2}\right)^{n_x} \geq \frac{2}{m+1} \left(\frac{x}{a_0}\right)^{\beta(m)}.$$

From this, the upper bound follows with $c_2 = -\frac{2}{m+1} a_0^{-\beta(m)} \log \varepsilon_0$ for all large x . By decreasing c_2 if necessary, one can of course get the desired inequality for all $x \geq 1$.

We now move to the lower bound. This time, choose $a_0 > 0$ so that $L(a_0) = \frac{1}{2} - \frac{\varepsilon_0}{2}$, and given $x \geq 1$, let $n_x := \lfloor \log_\gamma \frac{x}{a_0} \rfloor$. We again restrict to large x for which $n_x \geq 1$. This immediately yields

$$a_0 \geq x\gamma^{-n_x}.$$

Using (7) and monotonicity of L and h , we have

$$L(x) = h^{(n_x)}(L(x\gamma^{-n_x})) \leq h^{(n_x)}(L(a_0)).$$

By Lemma 12, we have that

$$\mathbb{P}(W_m \geq x) = \frac{1}{2} - L(x) \geq \frac{1}{2} - h^{(n_x)}(L(a_0)) = \frac{1}{2} - h^{(n_x)}\left(\frac{1}{2} - \frac{\varepsilon_0}{2}\right) \geq \left(\frac{\varepsilon_0}{2}\right)^{\binom{m+1}{2}n_x}.$$

An easy computation shows that one has, for all x for which $n_x \geq 1$, that

$$\left(\frac{m+1}{2}\right)^{n_x} \leq \frac{m+1}{2} \left(\frac{x}{a_0}\right)^{\beta(m)}.$$

From this, the lower bound follows for some c_1 for all large x , and by increasing c_1 if necessary, one can of course get the desired inequality for all $x \geq 1$. This proves part (b).

3.3. Proof of part (c)

We set out to show that $\beta(m)$ is strictly increasing. Since

$$\beta(m+2) - \beta(m) = \frac{\log \frac{m+3}{2} \log \gamma(m) - \log \frac{m+1}{2} \log \gamma(m+2)}{\log \gamma(m) \log \gamma(m+2)}, \tag{13}$$

it will suffice to show that the numerator in the right-hand side of (13) is strictly positive. Using the recursive structure of $\gamma(m)$, i.e., that $\gamma(m+2) = \frac{m+2}{m+1}\gamma(m)$, we aim to show that

$$\log \gamma(m) \log \frac{m+3}{m+1} - \log \frac{m+1}{2} \log \frac{m+2}{m+1} > 0.$$

For $x \geq 0$ a Taylor estimate for $\log(1+x)$ gives the lower and upper bounds $x - \frac{x^2}{2}$ and x , respectively. A lower bound on the numerator in the right-hand side of (13) is thus given by

$$\log \gamma(m) \left[\frac{2}{m+1} - \frac{2}{(m+1)^2} \right] - \frac{1}{m+1} \log \frac{m+1}{2}. \tag{14}$$

A lower bound on $\gamma(m)$ can be obtained from known bounds on the central binomial coefficient. For instance, Wallis' product formula states that $a_n := \prod_{k=1}^n \frac{2k}{2k-1} \frac{2k}{2k+1}$ converges to $\frac{\pi}{2}$ as $n \rightarrow \infty$. Since a_n is increasing we have $a_n \leq \frac{\pi}{2}$ for all $n \geq 1$, leading to the bound

$$\binom{2n}{n} \geq 4^n \sqrt{\frac{2}{\pi(2n+1)}}.$$

Consequently $\gamma(m) \geq \sqrt{2m/\pi}$ for all $m \geq 3$. After multiplication by $m+1$, a lower bound on the expression in (14) is given by

$$\log \frac{2m}{\pi} \left[1 - \frac{1}{m+1} \right] - \log \frac{m+1}{2} = \log \frac{4}{\pi} - \log \frac{m+1}{m} - \frac{1}{m+1} \log \frac{2m}{\pi}.$$

One may check that the latter expression is increasing in m and positive for $m = 13$. Using the slightly sharper lower bound $\frac{4^n}{\sqrt{\pi n}}(1 - \frac{1}{8n})$ on the central binomial coefficient, obtained from the Stirling series, one may arrive at an alternative lower bound on the difference in (13), which is positive for all $m \geq 5$. In either case, one further checks that $\beta(m) < \beta(m+2)$ for the remaining values of m by hand, so that $\beta(m)$ is strictly increasing for all $m \geq 3$.

3.4. Proof of part (d)

Since m will now be changing, it is natural to now write L_m instead of L . The distribution given by $F_m(x) = \frac{1}{2} + L_m(x)$ has density $L'_m(x)$. We will show that $\lim_{m \rightarrow \infty} L'_m(x) = e^{-\pi x^2}$, which we recognize as the density of a centered normal distribution with variance $1/(2\pi)$. By virtue of Scheffé’s theorem (see e.g. [9]), pointwise convergence of densities implies the desired weak convergence of F_m to a normal distribution. By symmetry, it suffices to prove this for $x \geq 0$ which we now assume to be the case.

We first show that $L'_m(x\gamma^{-1}) \rightarrow 1$ as $m \rightarrow \infty$, for every $x \geq 0$. First recall that by Proposition 10, we have that $L'_m(x) \leq 1$ for all $x \geq 0$. Using (12) and (5) and then that $L'_m \leq 1$, we find that

$$L'_m(x\gamma^{-1}) = \prod_{k=1}^{\infty} [1 - 4[L_m(x\gamma^{-k-1})]^2]^{\frac{m-1}{2}} \geq \prod_{k=1}^{\infty} [1 - 4x^2\gamma^{-2(k+1)}]^{\frac{m-1}{2}}.$$

Since $e^{-2y} \leq 1 - y$ for small positive y , replacing the terms of the product by an exponential, we obtain that for all large m

$$\begin{aligned} L'_m(x\gamma^{-1}) &\geq \prod_{k=1}^{\infty} \exp(-4x^2(m-1)\gamma^{-2(k+1)}) = \exp\left(-4x^2(m-1) \sum_{k=1}^{\infty} \gamma^{-2(k+1)}\right) \\ &\geq \exp(-8x^2m\gamma^{-4}), \end{aligned}$$

where we in the last step have used that $\gamma^2 \geq 2$. Since $\gamma = \gamma(m)$ increases at the rate of \sqrt{m} , we may conclude that $L'_m(x\gamma^{-1}) \rightarrow 1$ as $m \rightarrow \infty$. Moreover, the convergence is uniform in x over compact sets. Consequently, for every $\varepsilon > 0$ and $x \geq 0$ there is m_0 such that $L_m(x\gamma^{-1}) \geq (1 - \varepsilon)x\gamma^{-1}$ for all $m \geq m_0$.

Second, again using (12) and (5), or differentiating (7), we arrive at

$$L'_m(x) = L'_m(x\gamma^{-1})[1 - 4[L_m(x\gamma^{-1})]^2]^{\frac{m-1}{2}}.$$

Together with our previous conclusions we find that

$$L'_m(x\gamma^{-1})[1 - 4x^2\gamma^{-2}]^{\frac{m-1}{2}} \leq L'_m(x) \leq L'_m(x\gamma^{-1})[1 - (1 - \varepsilon)^2 4x^2\gamma^{-2}]^{\frac{m-1}{2}}.$$

Taking limits, first as $m \rightarrow \infty$ and then as $\varepsilon \rightarrow 0$, leaves us with

$$\lim_{m \rightarrow \infty} L'_m(x) = e^{-\pi x^2},$$

as required.

4. All measures are distributional limits

In this section, we prove Theorem 3 and Propositions 4 and 5.

Proof of Theorem 3. The main part of the proof will be to, in both settings (a) and (b), prove the result for a restricted class of probability measures, namely those μ of the form $\sum_{i=1}^k q_i \delta_{x_i}$, i.e., having finite support. To see why this will suffice in order to obtain the general result, we first state a simple lemma, whose proof is left to the reader, concerning metric spaces. Assume, in a metric space, we are given x_m converging to x_∞ and for each m , we have $x_{m,n}$ converging to x_m as $n \rightarrow \infty$. Then there is a sequence m_n (not necessarily strictly increasing) so that we have that $x_{m_n,n}$ converges to x_∞ as $n \rightarrow \infty$.

We note that it is well known that convergence in distribution is metrizable. Assume now that we are given an arbitrary probability measure μ and a sequence $(a_n)_{n \geq 1}$ satisfying the stated properties. It is clear we can find a sequence $(\mu_m)_{m \geq 1}$, each with finite support as above, converging to μ . Assume that, for each m , we can find a sequence of Boolean functions $(f_{m,n})_{n \geq 1}$ such that $f_{m,n}$ is defined on n bits and $a_n(T(f_{m,n}) - \frac{1}{2})$ approaches, as

$n \rightarrow \infty, \mu_n$ in distribution. By the above general metric space result, there exists a sequence $(m_n)_{n \geq 1}$ (not necessarily strictly increasing) so that $a_n(T(f_{m_n, n}) - \frac{1}{2})$ approaches μ in distribution, as $n \rightarrow \infty$. This shows that it will suffice to prove parts (a) and (b) for measures μ having finite support.

Proof of part (a). Assume that a probability measure μ of the form $\sum_{i=1}^k q_i \delta_{x_i}$ is given and let $(a_n)_{n \geq 1}$ satisfy $1 \ll a_n \ll \sqrt{n}$. We may assume that $x_1 < x_2 < \dots < x_k$ and that the q_i 's are all positive. For each $i = 1, 2, \dots, k$ fix $y_i \in \mathbb{R}$ such that

$$1 - \Phi(y_i) = q_1 + q_2 + \dots + q_i,$$

where $\Phi(\cdot)$ denotes the distribution function of the standard Gaussian. (Of course, this defines y_k to be $-\infty$, but we allow this slight abuse of notation.)

Now let E_i denote the event that the proportion of 1's among the n bits is at least $\frac{1}{2} + x_i/a_n$, and let F_i denote the event that the proportion of 1's among the first $\lfloor a_n \rfloor$ bits is at least $\frac{1}{2} + y_i/(2\sqrt{a_n})$. Although not explicit in the notation, these events depend on n . Notice further that the events are defined so that $E_1 \supseteq E_2 \supseteq \dots \supseteq E_k$ and $F_1 \subseteq F_2 \subseteq \dots \subseteq F_k$ hold. Finally, we define f_n as the indicator function of the event $\bigcup_{i=1}^k (E_i \cap F_i)$.

Let $p_n = \frac{1}{2} + x/a_n$. To complete the proof of part (a) we need to verify that

$$\mathbb{P}\left(a_n\left(T_n - \frac{1}{2}\right) \leq x\right) = \mathbb{P}_{p_n}(E_i \cap F_i \text{ for some } i)$$

tends to 0, $\sum_{i=1}^j q_i$ or 1, depending on whether $x < x_1$, $x \in (x_j, x_{j+1})$ and $j = 1, 2, \dots, k - 1$, or $x > x_k$. We first examine the events E_i and F_i . Appealing to the Lindeberg-Feller central limit theorem, or Chebyshev's inequality, we find that

$$\mathbb{P}_{p_n}(E_i) = \mathbb{P}_{p_n}\left(\frac{1}{n} \sum_{j \in [n]} \omega_j \geq \frac{1}{2} + \frac{x_i}{a_n}\right) \rightarrow \begin{cases} 0, & x < x_i, \\ 1, & x > x_i, \end{cases} \tag{15}$$

as by assumption $a_n \ll \sqrt{n}$. Moreover, as $a_n \gg 1$, using Lindeberg-Feller, for any $x \in \mathbb{R}$

$$\mathbb{P}_{p_n}(F_i) = \mathbb{P}_{p_n}\left(\frac{1}{\lfloor a_n \rfloor} \sum_{j=1}^{\lfloor a_n \rfloor} \omega_j \geq \frac{1}{2} + \frac{y_i}{2\sqrt{a_n}}\right) \rightarrow 1 - \Phi(y_i). \tag{16}$$

(The above abuse of notation is here manifested in that F_k equals the whole sample space.)

Since $\mathbb{P}_{p_n}(E_i \cap F_i \text{ for some } i)$ is at most $\mathbb{P}_{p_n}(E_i \text{ for some } i)$, the case $x < x_1$ is immediate from (15). Similarly, as F_k equals the whole sample space, $\mathbb{P}_{p_n}(E_k)$ gives a lower bound on $\mathbb{P}_{p_n}(E_i \cap F_i \text{ for some } i)$, so also the case $x > x_k$ is immediate from (15). For $j = 1, 2, \dots, k - 1$ and $x \in (x_j, x_{j+1})$, using (15), (16) and the fact that $F_1 \subseteq F_2 \subseteq \dots \subseteq F_k$, gives

$$\lim_{n \rightarrow \infty} \mathbb{P}_{p_n}(E_i \cap F_i \text{ for some } i) = \lim_{n \rightarrow \infty} \mathbb{P}_{p_n}(F_j) = q_1 + q_2 + \dots + q_j, \tag{17}$$

due to the definition of the y_i 's, as required.

Proof of part (b). Assume again that a probability measure μ of the form $\sum_{i=1}^k q_i \delta_{x_i}$ is given and let $(a_n)_{n \geq 1}$ satisfy $\log n \ll a_n \ll \sqrt{n}$. We may assume that $x_1 < x_2 < \dots < x_k$ and that the q_i 's are all positive. As before, let E_i denote the event that the proportion of 1's among the n bits is at least $\frac{1}{2} + x_i/a_n$.

Let $\ell_n = \lfloor 2 \log_2 n \rfloor$. In order to define events F_i we extend the partial ordering on binary strings of length ℓ_n , in which $y \geq y'$ if y dominates y' coordinate-wise, to a total ordering. (One such ordering is inherited from the set of integers through their binary representation.) Consider the n bits of ω positioned in a circle and, for each string $y \in \{0, 1\}^{\ell_n}$, let $F(y)$ denote the event that ω contains an interval of length ℓ_n on which ω is at least as large as y with respect to the total ordering. We claim that it is possible to choose $y_1^n \geq y_2^n \geq \dots \geq y_k^n$ in $\{0, 1\}^{\ell_n}$ such that $F_i := F(y_i^n)$, for every $i = 1, 2, \dots, k$, $x \in \mathbb{R}$ and with $p_n = \frac{1}{2} + x/a_n$, satisfies

$$\lim_{n \rightarrow \infty} \mathbb{P}_{p_n}(F_i) = q_1 + q_2 + \dots + q_i. \tag{18}$$

These definitions guarantee that $E_1 \supseteq E_2 \supseteq \dots \supseteq E_k$ and $F_1 \subseteq F_2 \subseteq \dots \subseteq F_k$ hold also here.

Assuming that such a selection of y_i^n 's can be made, we define f_n as the indicator function of the event $\bigcup_{i=1}^k (E_i \cap F_i)$. The resulting function is monotone (with respect to the partial ordering on $\{0, 1\}^n$) and invariant with respect to rotations of the bits and therefore transitive.

Before proving the claim we argue for why this would complete the proof of part (b). As in part (a), to complete the proof we need to verify that

$$\mathbb{P}\left(a_n\left(T_n - \frac{1}{2}\right) \leq x\right) = \mathbb{P}_{p_n}(E_i \cap F_i \text{ for some } i)$$

tends to 0, $\sum_{i=1}^j q_i$ or 1, depending on whether $x < x_1$, $x \in (x_j, x_{j+1})$ and $j = 1, 2, \dots, k - 1$, or $x > x_k$. As the events E_i are defined just as above, and the monotone structure $F_1 \subseteq F_2 \subseteq \dots \subseteq F_k$ holds also here, these conclusions would follow from the claim as in part (a).

It remains to prove the claim. We first show that for every $n \geq 1$ and $q \in [0, 1]$ there exists $y \in \{0, 1\}^{\ell_n}$ such that $|\mathbb{P}_{1/2}(F(y)) - q| \leq 2/n$. Let $y_1 = \max\{y : \mathbb{P}_{1/2}(F(y)) \geq q\}$ and denote its successor in the total ordering by y_2 , should it exist (otherwise set $F(y_2) = \emptyset$). Then $\mathbb{P}_{1/2}(F(y_2)) < q$ and we have

$$0 \leq \mathbb{P}_{1/2}(F(y_1)) - \mathbb{P}_{1/2}(F(y_2)) \leq \mathbb{P}_{1/2}(\text{some interval of } \omega \text{ equals } y_1) \leq n2^{-\ell_n} \leq 2/n,$$

where the second-to-last inequality comes from a first moment estimate. In particular, for $y = y_1$ we obtain $|\mathbb{P}_{1/2}(F(y)) - q| \leq 2/n$.

We next show that if some (sequence of) $y \in \{0, 1\}^{\ell_n}$ satisfies $\mathbb{P}_{1/2}(F(y)) \rightarrow q$ as $n \rightarrow \infty$, then also $\mathbb{P}_{p_n}(F(y)) \rightarrow q$. Recall the monotone coupling $(\eta_p)_{p \in [0, 1]}$ of elements in $\{0, 1\}^n$. Assume that $x \geq 0$ so that $p_n \geq 1/2$, and let A denote the event that $\eta_{1/2}$ contains no interval on which it is at least as large as y (in the total ordering) but that η_{p_n} does. Then,

$$\mathbb{P}_{p_n}(F(y)) - \mathbb{P}_{1/2}(F(y)) = \mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A|\eta_{p_n})] \leq \ell_n \left(p_n - \frac{1}{2}\right) / p_n \leq 2x\ell_n/a_n, \tag{19}$$

since $\mathbb{P}(A|\eta_{p_n})$ either equals zero, in the case η_{p_n} does not contain an interval on which it is at least as large as y , or is bounded by $\ell_n(p_n - \frac{1}{2})/p_n$, since if such an interval exists then at least one bit in this interval (or the first in some ordering, if there are several intervals with this property) must have changed its value as p ranges from $\frac{1}{2}$ to p_n . Since $a_n \gg \log n$, the upper bound in (19) tends to zero as $n \rightarrow \infty$, as claimed. The case $x < 0$ is analogous.

This proves the claim that we may choose y_i^n 's such that (18) holds for all $x \in \mathbb{R}$, and thus completes the proof of part (b). □

Remark. In order to obtain a sequence $(f_n)_{n \geq 1}$ of monotone graph properties on n vertices and $\binom{n}{2}$ edges (which are the variables here), with the property described in the remark following Theorem 3, only minor modifications to the above construction are necessary: Define the E_i 's as before, but now on $\binom{n}{2}$ bits. Let $\ell_n = \lfloor 4 \log_2 n \rfloor$, say, and extend the usual partial ordering on the set of unlabeled graphs on ℓ_n vertices to a total ordering. Given an unlabeled graph y on ℓ_n vertices, define $F(y)$ as the event that ω contains an induced subgraph at least as large as y with respect to the total ordering. The calculations needed to verify that one may choose graphs y_i^n so that (18) holds also in this setting are straightforward.

We continue investigating the possible behavior of the scaling coefficients.

Proof of Proposition 4. Let μ be a non-degenerate probability measure and $(f_n)_{n \geq 1}$ some sequence of monotone Boolean functions for which $a_n(T_n - b_n)$ approaches μ in distribution. Denote by F the distribution function associated to μ .

To prove part (a), we first recall the well known fact that for monotone Boolean functions the total influence, for p bounded away from 0 and 1, is of order at most \sqrt{n} . The Margulis-Russo formula then implies that there is an upper bound of order \sqrt{n} on the derivative of $\mathbb{P}_p(f_n = 1)$, for p bounded away from 0 and 1. Now, pick $x_1 < x_2$ at which F is continuous and satisfies $F(x_1) \leq 1/3$ and $F(x_2) \geq 2/3$. With $p_i = b_n + x_i/a_n$, we thus obtain, for all large n ,

$$\frac{a_n}{4(x_2 - x_1)} \leq \frac{\mathbb{P}_{p_2}(f_n = 1) - \mathbb{P}_{p_1}(f_n = 1)}{p_2 - p_1} \leq \frac{a_n}{x_2 - x_1}. \tag{20}$$

Via the mean value theorem, the first inequality in (20) gives a lower bound on the derivative of $\mathbb{P}_p(f_n = 1)$ for some $p \in [p_1, p_2]$, which, consequently, yields an upper bound on a_n of order \sqrt{n} .

The proof of part (b) is similar to that of part (a). Recall the well known fact from KKL [31] and its extensions that the total influence of transitive Boolean functions grows at least of order $\log n$ when their variance is bounded away from 0 and 1. With the mean value theorem, the Margulis-Russo formula and the second inequality in (20) we thus obtain a corresponding lower bound on a_n , of order $\log n$.

The proofs for parts (c), (d) and (e) are also similar to each other. Let, again, $x_1 < x_2$ be continuity points of F and set $p_i = b_n + x_i/a_n$. Using the mean value theorem we find $q_n \in [p_1, p_2]$ so that

$$\frac{F(x_2) - F(x_1)}{x_2 - x_1} = \lim_{n \rightarrow \infty} \frac{\mathbb{P}_{p_2}(f_n = 1) - \mathbb{P}_{p_1}(f_n = 1)}{x_2 - x_1} = \lim_{n \rightarrow \infty} \frac{1}{a_n} \frac{d}{dp} \mathbb{P}_p(f_n = 1) \Big|_{q_n}. \tag{21}$$

For (c) we assume that $(a_n)_{n \geq 1}$ is bounded above and that $0 < F(x_1) \leq F(x_2) < 1$. Recall the discrete Poincaré inequality which states that for any Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$

$$\sum_{i=1}^n \text{Inf}_i^p(f) \geq \text{Var}_p(f). \tag{22}$$

Hence, a lower bound on (21) of $\liminf_{n \rightarrow \infty} \text{Var}_{q_n}(f_n) / (\sup_{n \geq 1} a_n)$, which is strictly positive, is obtained via the Margulis-Russo formula. Consequently $F(x_1) < F(x_2)$, from which (c) follows.

For transitive functions KKL [31] and its extensions improve on the lower bound in (22) with a factor of order $\log n$. In this setting, assuming that $(a_n / \log n)_{n \geq 1}$ is bounded above, we thus obtain a lower bound on (21) of $\liminf_{n \rightarrow \infty} \text{Var}_{q_n}(f_n) (\inf_{n \geq 1} \log n / a_n)$, which is strictly positive by assumption. Again we find that $F(x_1) < F(x_2)$, which settles part (d).

For (e) we assume that a_n / \sqrt{n} is bounded away from 0 and use the upper bound of order \sqrt{n} on the derivative of $\mathbb{P}_p(f_n = 1)$, for p bounded away from 0 and 1, to obtain a uniform upper bound in (21), showing that F is Lipschitz continuous. \square

Remark. While the above proof of part (c) shows that no monotone Boolean function f may have $\mathbb{P}_p(f = 1)$ making a finite number of ‘‘jumps’’ but otherwise remaining more or less constant, it is still possible to have sudden jumps between which $\mathbb{P}_p(f = 1)$ grows linearly. An example would be the event A consisting of all configurations for which either $\omega_1 = 1$ and the proportion of 1’s is at least $1/3$ or the proportion of 1’s is at least $2/3$.

As mentioned in the introduction, one easily shows that no subsequence of the probability measures giving equal weight to the points in $\Omega_m = \{\pm 2^k : k = 1, 2, \dots, m\}$ can be normalized in order to give a nondegenerate limit. We base the proof of Proposition 5 on this example.

Proof of Proposition 5. We prefer to work with continuous distributions. Let μ_m be the measure whose density function equals $\frac{1}{2^m}$ for $x \in [k - \frac{1}{2}, k + \frac{1}{2}]$ and $k \in \Omega_m$, and 0 otherwise; let F_m denote the corresponding distribution function. Then F_m is continuous and μ_m effectively has the same properties as the uniform measure on Ω_m . According to Theorem 3 we may choose $a_n = n^{1/4}$, say, and monotone Boolean functions $f_{m,n} : \{0, 1\}^n \rightarrow \{0, 1\}$ such that $a_n(T(f_{m,n}) - \frac{1}{2})$ tends to μ_m in distribution, as $n \rightarrow \infty$. Writing $F_{m,n}$ for the distribution function of $a_n(T(f_{m,n}) - \frac{1}{2})$ and using that F_m is continuous, we find for each m an integer n_m such that

$$\sup_{x \in \mathbb{R}} |F_{m,n}(x) - F_m(x)| \leq \frac{1}{m} \quad \text{for all } n \geq n_m. \tag{23}$$

Define $f_n := f_{m,n}$ for $n \in [n_m, n_{m+1})$, and note that f_n is a monotone function on n variables.

Let $m_n := \max\{m \in \mathbb{N} : n_m \leq n\}$. Now, assume there are nonnegative sequences $(b_n)_{n \geq 1}$ and $(c_n)_{n \geq 1}$, and a nondegenerate probability measure with distribution function F such that, along some subsequence, $F_{m_n,n}(c_n x + b_n) \rightarrow F(x)$ for all continuity points of F . Then,

$$|F_{m_n}(c_n x + b_n) - F(x)| \leq |F_{m_n}(c_n x + b_n) - F_{m_n,n}(c_n x + b_n)| + |F_{m_n,n}(c_n x + b_n) - F(x)|,$$

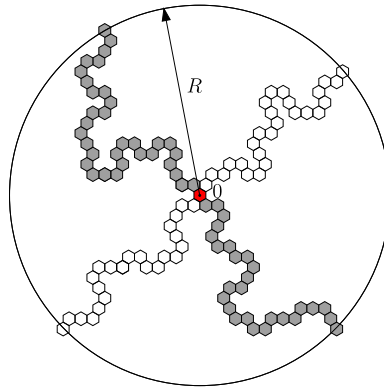


Fig. 1. A realization of the *four-arm* event.

which, for continuity points of F , would tend to zero along this subsequence, in virtue of (23). This would contradict the fact that no subsequence of μ_m can be normalized to obtain a nondegenerate limit, and therefore shows that no subsequence of $(T(f_n))_{n \geq 1}$ can be normalized to obtain a nondegenerate limit. \square

Appendix: The tail of the crossing probability in near-critical percolation (Authored by Gábor Pete)

The goals of this appendix are to prove Theorem 2 of the main text, to draw attention to an interesting difference between the flip times of crossing events in near-critical and dynamical percolation in the plane, and to examine, on an intuitive level, how near-critical high-dimensional percolation may behave.

A.1. Planar percolation

We will work with site percolation on the triangular lattice \mathbb{T} (which is the same as coloring the faces of the hexagonal tiling) with mesh size 1, at density close to the critical value $p_c = 1/2$. See [19,53] for background. Let $\text{LR}_{\mathcal{Q}}$ denote the left-to-right crossing event in a nice quad \mathcal{Q} , by which we mean the image of the square $[0, 1]^2$ under a smooth injective map into \mathbb{C} . If we magnify \mathcal{Q} by a factor of ρ , the new quad will be denoted by $\rho\mathcal{Q}$, the center of magnification being irrelevant. Furthermore, let $\alpha_4(n)$ denote the critical alternating four-arm probability from a given site to Euclidean distance n , and let $r(n) := 1/(n^2\alpha_4(n))$. The asymptotics $r(n) = n^{-3/4+o(1)}$ was proved in [51].

Consider now the canonical coupling of percolation configurations at different densities: take $U_x \sim \text{Unif}[0, 1]$ i.i.d. for all vertices $x \in \mathbb{T}$, and for any $\lambda \in (-\infty, \infty)$ and n large enough so that $\lambda r(n) \in [-1/2, 1/2]$, define $\omega_n(\lambda)$ to be the configuration where x is open iff $U_x \leq 1/2 + \lambda r(n)$. This process $\{\omega_n(\lambda)\}_{\lambda \in \mathbb{R}}$ is called the near-critical ensemble. Note that the flip time T_n for $\text{LR}_{[0,n]^2}$ considered in the main text, with rescaling $W_n := n^2\alpha_4(n)(T_n - 1/2)$, exactly satisfies $\{W_n \leq \lambda\} = \{\omega_n(\lambda) \in \text{LR}_{[0,n]^2}\}$.

It is proved in [15,16] that the process $\{\omega_n(\lambda)\}_{\lambda \in \mathbb{R}}$ has a scaling limit in a certain topology: for any fixed λ , the static topology encodes quad-crossings, while as a process, the Skorokhod topology of càdlàg processes is used; see [16, Theorem 1.5] for the precise statement. (The time-parametrization of $\omega_n(\lambda)$ in [16] is slightly different from the above definition, but this makes no real difference.) It is also proved in [16, Proposition 9.6] that this limiting càdlàg process has no jumps at deterministic values of λ , which implies that for any $\lambda \in \mathbb{R}$, the limit

$$f(\lambda, \mathcal{Q}) := \lim_{n \rightarrow \infty} \mathbf{P}_{p_c + \lambda r(n)}[\text{LR}_n \mathcal{Q}] \quad (24)$$

exists for any nice quad \mathcal{Q} (which does not automatically follow from the existence of the limit as a process, because of the Skorokhod topology of càdlàg processes). Moreover, this limit is absolutely continuous in λ , and most importantly, it is non-trivial: it satisfies $f(\lambda, \mathcal{Q}) \in (0, 1)$, and

$$\lim_{\lambda \rightarrow -\infty} f(\lambda, \mathcal{Q}) = 0 \quad \text{and} \quad \lim_{\lambda \rightarrow \infty} f(\lambda, \mathcal{Q}) = 1. \quad (25)$$

In fact, these properties were already known from Kesten's work [34,43], for any subsequential limit, at that time. Briefly, in the entire critical window where $\mathbf{P}_p[\text{LR}_{n\mathcal{Q}}] \in (\varepsilon, 1 - \varepsilon)$ holds, the expected number of pivotals is comparable to the scaling factor $n^2\alpha_4(n)$ (with constant factors depending on \mathcal{Q} and ε), which implies absolute continuity and (25) using Russo's formula.

The above results imply Part (a) of Theorem 2.

For the proof of Part (b), we will also need that the limit $f(\lambda, \mathcal{Q})$ in (24) is conformally covariant, proved in [16, Theorem 10.3]. Instead of defining exactly what this means, let us just give a special case that we will use:

$$f(\rho\lambda, \mathcal{Q}) = f(\lambda, \rho^{4/3}\mathcal{Q}), \quad (26)$$

for any scaling factor $\rho > 0$. For simplicity, we will just take $\mathcal{Q} = [0, 1]^2$. Then, the limiting tail behaviour of the rescaled flip time W_n is determined by the following result, proved below, together with the well-known duality $f(-\lambda, [0, 1]^2) = 1 - f(\lambda, [0, 1]^2)$:

Theorem 13. *As $\lambda \rightarrow \infty$, we have the superexponential decay*

$$f(-\lambda, [0, 1]^2) = \exp(-\Theta(\lambda^{4/3})),$$

where, as usually, $g(\lambda) = \Theta(h(\lambda))$ means the existence of universal constants $0 < c < C < \infty$ such that $c < g(\lambda)/h(\lambda) < C$ holds for all λ in question.

Besides the question raised in the main text, another motivation for Theorem 13 is [20], where the analogous tail behaviour was studied for the scaling limit of dynamical percolation. Namely, if we start with critical percolation, then resample each site at rate $r(n)$, keeping the configuration stationary, then we may look at

$$g(t, \mathcal{Q}) := \lim_{n \rightarrow \infty} \mathbf{P}[\text{LR}_{n\mathcal{Q}} \text{ does not hold at any moment in } [0, t]]. \quad (27)$$

Again, this limit exists and is conformally covariant by [15,16]. Then, regarding the tail behaviour, it was proved in [20] using general Markov chain arguments such as spectral computations and a dynamical (space-time) FKG-inequality, that there exists an absolute constant $c > 0$, and for every $K > 0$ some $c_K > 0$, such that

$$\exp(-ct) \leq g(t, [0, 1]^2) \leq c_K t^{-K}, \quad (28)$$

for all $t \geq 1$. Furthermore, the present author was speculating in [45], using non-rigorous renormalization ideas (motivated by [38,39,48]) and a very strong universality hypothesis, that the true behaviour could be $\exp(-t^{2/3+o(1)})$. Several people in the community agreed that this speculation looked quite solid (even if non-rigorous) as a lower bound, while more questionable as an upper bound. And, as typical for these planar percolation scaling limits, that argument seemed to be working equally well for the symmetric (dynamical) and asymmetric (near-critical) versions. However, our present Theorem 13 violates not only this bold prediction for the near-critical case, but it also shows that $f(t, \mathcal{Q})$ does not satisfy the rigorous exponential lower bound of (28) for $g(t, \mathcal{Q})$, hence this tail probability question turns out to be an instance where the asymmetric versus symmetric dynamical versions of critical percolation show drastically different behaviour. Regarding the true decay in the symmetric dynamical version, our simulations suggest a subexponential decay, but are far from being conclusive, and are even further from giving a prediction for the exponent. See Figure 2.

The proof of Theorem 13 is very simple, given the results of [15,16] cited above, and standard percolation techniques proving exponential decay for certain connection probabilities, as can be found in [43, Section 7]. This is somewhat similar to [8], where Duminil-Copin showed, again building on [15,16], that the super-critical percolation Wulff shape is asymptotically circular, as the density p approaches p_c .

Proof of Theorem 13. By the scaling covariance (26), we need to show that

$$f(-1, [0, \lambda^{4/3}]^2) = \exp(-\Theta(\lambda^{4/3})), \quad (29)$$

as $\lambda \rightarrow \infty$.

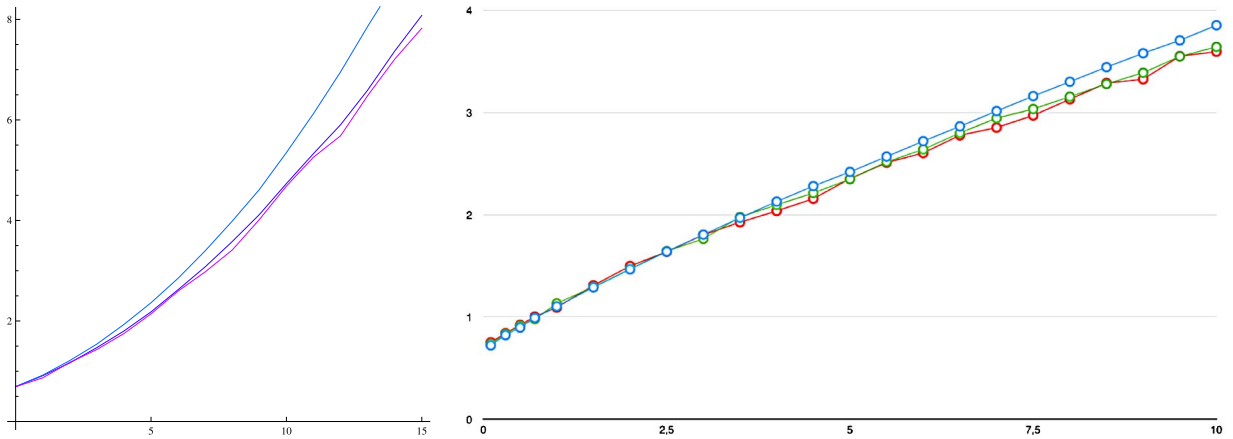


Fig. 2. On the left, simulation results are shown for $-\log f(\lambda, [0, 1]^2)$, with the near-critical percolation parameter varying from $\lambda = 0$ to 1.5, board sizes $n = 10, 100, 500$. On the right, simulation results are shown for $-\log g(t, [0, 1]^2)$ in dynamical percolation for scaled time going from $t = 0$ to 10, on board sizes $n = 10, 100, 200$. In both cases, the values are lower and have more fluctuations as n increases, since fewer simulation runs were feasible. The superexponential decay for $f(\lambda, [0, 1]^2)$ is apparent, the subexponential decay for $g(t, [0, 1]^2)$ is less so.

By [43, Lemma 39 and (7.28)], there exist $0 < C_1 \leq C_2 < \infty$ such that for all N and $p < p_c$,

$$C_1 \exp(-C_2 N/L(p)) \leq \mathbf{P}_p[\mathbf{LR}_{[0, N]^2}] \leq C_2 \exp(-C_1 N/L(p)), \tag{30}$$

where $L(p)$ is a correlation length; say,

$$L(p) := \inf\{n : \mathbf{P}_p[\mathbf{LR}_{[0, n]^2}] < 1/100\}. \tag{31}$$

Let us remark that (30) is nothing mysterious: the main reason for it is that $1/100$ is small enough so that a variation of the Peierls contour argument works on a renormalized lattice with mesh $L(p)$.

On the other hand, it is well-known (see, e.g., Proposition 34 in [43]) that we have $L(p_c - r(n)) = \Theta(n)$, with the constant factors in Θ depending, of course, on $1/100$ in definition (31). Now take $p = p_c - r(n)$ and $N = \lambda^{4/3}n$ in (30), then send $n \rightarrow \infty$ to see that (29) holds. \square

A.2. What about high dimensions?

It is proved in Theorem 3 of the main text that for any probability distribution function $F(\cdot)$ there exists a sequence of monotone Boolean functions $\{f_n\}$ and some parameters p_n, b_n such that $\mathbf{P}_{p_n + \lambda b_n}[f_n = 1] \rightarrow F(\lambda)$ as $n \rightarrow \infty$, for all $\lambda \in \mathbb{R}$. However, all “natural” examples of limit distributions $F(\cdot)$ found so far have exponential or superexponential tails. The exponent $4/3$ in the superexponential decay $\exp(-|\lambda|^{4/3})$ of the previous section appears to be a rather direct consequence of the planar correlation length exponent being $\nu = 4/3$ (i.e., the correlation length (31) satisfies $L(p) = |p - p_c|^{-4/3 + o(1)}$ as $p \nearrow p_c$) and of the critical window being $r(n) = n^{-3/4 + o(1)}$. Therefore, one might hope (as the present author did in the first version of this appendix) that crossing events in near-critical percolation on \mathbb{Z}^d , where d is high enough so that already mean-field behaviour takes place, with $\nu = 1/2 < 1$, could have subexponential tails. However, as we will briefly explain, this turns out to be very naive, and we expect now that the lower tail in high dimension is exponential in $|\lambda|$, while the upper tail is doubly exponential, similarly to some of the examples in Section 2 of the main text.

By high-dimensional percolation we will mean d high enough so that the critical two-point connectivity function already scales like Green’s function for simple random walk:

$$\mathbf{P}_{p_c}[\underline{x} \longleftrightarrow \underline{y}] = \Theta(\|\underline{x} - \underline{y}\|^{2-d}). \tag{32}$$

This is conjectured to be the case for $d > 6$, proved for $d \geq 19$ in [22], and recently for $d \geq 11$ in [12]. From now on, we will always assume that this mean-field behavior holds. Two-point connectivity can also be used to define a

near-critical correlation length $\xi(p)$:

$$\mathbf{P}_p[\underline{x} \longleftrightarrow \underline{y} \not\leftrightarrow \infty] = \exp(-\Theta(\|\underline{x} - \underline{y}\|/\xi(p))),$$

which makes sense both for $p < p_c$ and $p > p_c$. In the planar case, it is known that $\xi(p) \asymp L(p)$, with $L(p)$ defined in (31); see [43, Theorem 33]. The mean-field value of the exponent in $\xi(p) = |p - p_c|^{-\nu+o(1)}$ is $\nu = 1/2$, proved for $p < p_c$ in [21]. In analogy with the 2-dimensional case, this may suggest that the critical window for left-to-right crossing in a box of side-length n is of size $n^{-2+o(1)}$. This will in fact align with our conjectures, but in a highly non-trivial way. For the same reasons, the near-critical upper and lower tails will behave very differently in high-dimensional percolation than in the planar case.

The first key difference from the planar case, observed in [1], is that at the critical density $p_c(\mathbb{Z}^d)$, with high probability there are already many disjoint left-to-right crossings in a large box $[n]^d$. Indeed, as a back-of-the-envelope calculation, using the critical two-point function (32), we get that the expected number of vertices y on the right side of $[n]^d$ connected to a given vertex x on the left is on the order of $n^{2-d}n^{d-1} = n$. This strongly suggests that the finite size critical density $p_c(n)$, where the probability of having a left-to-right crossing is exactly $1/2$, is below $p_c(\mathbb{Z}^d)$, and $p_c(\mathbb{Z}^d)$ is already outside of the critical window for left-to-right crossing.

For the precise behavior, we conjecture the following:

Conjecture 14. *For left-to-right crossing in percolation on $[n]^d$, $d > 6$, the critical density is*

$$p_c(n) = p_c(\mathbb{Z}^d) - (c_d + o(1)) \frac{\log n}{n^2},$$

with $c_d > 0$. The critical window is

$$p = p_c(n) + \frac{\lambda}{n^2}, \quad \lambda \in (-\infty, \infty).$$

In this window, the number of disjoint crossings should be approximately Poisson, with mean $\exp(\Theta(\lambda))$. Hence, for a very negative λ , the probability of a left-to-right crossing should be $\exp(-\Theta(|\lambda|))$, while, for a large positive λ , the probability of having no left-to-right crossing should be $\exp(-\exp(\Theta(\lambda)))$.

This location of the window is also conjectured, independently, by Gady Kozma [35], while a Gumbel limit distribution is confirmed by computer simulations of Eren Metin Elçi [10].

We will now explain how one may arrive at this conjecture. Then we will prove rigorously the upper bound $O(n^{-2/3})$ on the size of the critical window, partly due to Gady Kozma.

The heuristic picture of high-dimensional percolation, closely related to (32), is that the critical cluster is like a critical Galton-Watson tree, embedded into \mathbb{Z}^d as a branching random walk. A precise formulation of this is that the scaling limit of the Incipient Infinite Cluster is Integrated Infinite Canonical Super-Brownian Motion. This has been proved only very partially [23,28,29]; see [50] for a nice introduction into the limit object, and [26] for a survey. Therefore, in order to understand near-critical high-dimensional percolation, one could first look at near-critical Galton-Watson trees. One can show that, for any offspring distribution with mean $1 - \lambda\varepsilon$ and finite variance,

$$\mathbf{P}_{\text{GW}(1-\lambda\varepsilon)}[o \longleftrightarrow 1/\varepsilon] = \Theta(\varepsilon) \exp(-\Theta(\lambda)), \quad 0 \leq \lambda < 1/\varepsilon, \tag{33}$$

where $\{o \longleftrightarrow 1/\varepsilon\}$ denotes the event that the tree reaches generation $1/\varepsilon$. There are several ways to prove this estimate; one is to consider a depth-first type exploration process of the tree, and use martingale calculations for the resulting integer-valued random walk, as in [42]. Now, accepting the heuristic picture above, and using that a random walk path in \mathbb{Z}^d with $1/\varepsilon$ steps goes to a Euclidean distance about $\sqrt{1/\varepsilon}$ with large probability, the estimate (33) suggests that, for near-critical percolation on \mathbb{Z}^d ,

$$\mathbf{P}_{p_c-\lambda\varepsilon}[o \longleftrightarrow \partial B_{\sqrt{1/\varepsilon}}(o)] = \Theta(\varepsilon) \exp(-\Theta(\lambda)), \quad 0 \leq \lambda < p_c/\varepsilon, \tag{34}$$

where $B_r(o)$ denotes the ball of Euclidean radius r . Note here that, in principle, the cluster could reach to a large distance not only by the embedded critical tree having a large radius, but also by the embedding reaching unusually

far. However, changing the speed exponent $1/2$ has a probability cost that is exponential in the number of steps, $1/\varepsilon$, hence this strategy cannot improve on (34).

This was just heuristics, and in fact, only the critical case $\lambda = 0$ of (34) has been proved [37]: that is, the 1-arm exponent is 2. One approach to build an actual proof could be that the backbone of high-dimensional IIC is known to have a linear “chain of sausages” structure, with a linear number of pivotal edges, without too long sausages between them [28,29,36], hence the IIC must have a tree-like structure; also, its embedding into \mathbb{Z}^d is known to have 4-dimensional features [27,29]; these support the view that it is a large critical tree embedded as a branching random walk. Then, lowering the percolation density by $\lambda\varepsilon$, the probability of keeping the connection, which exists at criticality with probability $\Theta(\varepsilon)$, is at most the probability of not closing any of the pivotals, and should in fact be comparable to that: about $(1 - \lambda\varepsilon)^{\Theta(1/\varepsilon)} \asymp \exp(-\Theta(\lambda))$. This yields (34).

Now we would like to use (34) to understand the critical window of left-to-right crossing in a large box $[n]^d$. To start with, our previous calculation for the expected number of crossings at $p_c(\mathbb{Z}^d)$ can also be reproduced from the $\lambda = 0$ case of (34): the probability of x being connected to the right side is about n^{-2} , and conditioned on this event, the cluster should look like a conditioned critical branching random walk, having about n^4 vertices within Euclidean distance n , hence about n^3 vertices on the right side, altogether giving an expectation about $n^{-2}n^3 = n$. Furthermore, the expected number of disjoint clusters connecting the left and right sides is about $n^{d-1}n/n^{2\cdot 3} = n^{d-6}$, since there are $n^{d-1}n$ pairs of vertices x and y that are connected to each other, with each cluster having n^3 possible x 's and n^3 possible y 's. Indeed, it was proved in [1] that with probability tending to 1, there are at least order n^{d-6} disjoint connections, with the possibility of having some more spanning clusters that are thinner than 4-dimensional.

When, in order to find the critical window, we start decreasing p from $p_c(\mathbb{Z}^d)$, by (34), the expected number of disjoint crossings should be about

$$\mathbb{E}_{p_c - \frac{\lambda}{n^2}} [\#\{\text{disjoint crossings from left to right in } [n]^d\}] = \Theta(n^{d-6}) \exp(-\Theta(\lambda)); \tag{35}$$

the effect of the slight subcriticality on the size of the conditioned tree, and hence on the factor n^{d-6} should be only polynomial in λ , which is negligible compared to $\exp(-\lambda)$. We note that, to actually prove (35) or something a little weaker, instead of using (34), it might be easier to start directly from Aizenman’s proof [1].

Now, the finite size critical density $p_c(n)$ should be around a value $p_c - \lambda/n^2$ where the above expectation (35) is about 1, and the critical window should be where the expectation is independent of n . Within this window, by the approximate independence of the possible crossing clusters, we expect Poissonian behavior. These considerations give Conjecture 14.

We now turn to rigorous upper bounds on the window size.

Proposition 15. *For percolation on \mathbb{Z}^d with d high enough so that mean field behavior takes place, the critical window for left-to-right crossing in $[n]^d$ has width $O(n^{-2/3})$.*

Proof. As we mentioned above, the argument is partly due to Gady Kozma, and it will use a result of [44] on randomized algorithms computing Boolean functions.

Let $f_n : \{-1, 1\}^{[n]^d} \rightarrow \{-1, 1\}$ be the ± 1 -valued indicator function of the left-to-right crossing event in $[n]^d$, at a density p in the near-critical ε -window: $\mathbf{P}_p[f_n = 1] \in (\varepsilon, 1 - \varepsilon)$, for some fixed $\varepsilon > 0$. Note that, for n large enough, we have $1/(2d) < p < p_c(\mathbb{Z}^d)$.

Consider now the following randomized version of the algorithm used in [2, Section 4] to determine the value of f_n . Choose uniformly at random a coordinate hyperplane separating the left and right faces. Now explore the clusters of all the sites in this hyperplane to determine whether there is a crossing. This way, each bit (site or edge) of the box is explored with probability at most $O(n^{-2/3})$: either the bit is within distance $n^{1/3}$ of the chosen hyperplane, which has probability $n^{-2/3}$, or it will be queried only if in a cluster of radius at least $n^{1/3}$, which has probability $O(n^{-2/3})$, since the critical one-arm exponent is 2, and we are at density $p < p_c(\mathbb{Z}^d)$.

In the wording of [49], the above algorithm has revelation $O(n^{-2/3})$: the probability δ_i of revealing any bit i is $O(n^{-2/3})$. By the main inequality of [44], see also [18, Theorem XII.36], for the total influence we get

$$\text{Var}_p(f_n) \leq 4p(1 - p) \sum_{i \in [n]^d} \delta_i I_i^p(f_n) \leq O(n^{-2/3}) I^p(f_n). \tag{36}$$

Being in the ε -window, we have $4\varepsilon(1 - \varepsilon) \leq \text{Var}_p(f_n)$, hence (36) gives $c(\varepsilon)n^{2/3} \leq I^p(f_n)$, for some $c(\varepsilon) > 0$. By Russo’s formula, this implies that the near-critical ε -window for f_n has width at most $C(\varepsilon)n^{-2/3}$, as we claimed. \square

Let us mention that another way to do this argument, yielding a worse exponent, would have been via noise sensitivity and Fourier analysis; for background, see [18]. By [49, Theorem 1.8] (or more precisely, by its straightforward extension from density $p = 1/2$ to general p values bounded away from 0 and 1), the revelation $O(n^{-2/3})$ implies that f_n is sensitive to any noise $\gg n^{-1/3}$. In terms of the Fourier spectral sample $\text{Spec}^p(f_n)$, defined by

$$\mathbf{P}[\text{Spec}^p(f_n) = S] := \frac{\widehat{f}_n(S)^2}{\|f_n\|_2^2} = \widehat{f}_n(S)^2, \quad S \subset [n]^d,$$

where $\widehat{f}(S) := \mathbb{E}_p[f \chi_S^p]$ is the Fourier coefficient corresponding to the orthonormal basis

$$\chi_S^p(\omega) := \prod_{i \in S} \omega_i \left(\frac{1-p}{p} \right)^{\omega_i/2}, \quad \omega \in \{-1, 1\}^{[n]^d}$$

for $L^2(\{-1, 1\}^{[n]^d}, \mathbb{E}_p)$, this noise sensitivity can be written as follows: for any ε -window and any $\delta > 0$, if $\kappa > 0$ is small enough, then

$$\mathbf{P}[0 < |\text{Spec}^p(f_n)| < \kappa n^{1/3}] < \delta. \tag{37}$$

On the other hand, $\mathbf{P}[\text{Spec}^p(f_n) = \emptyset] = \mathbb{E}_p[f_n]^2 \leq (1 - 2\varepsilon)^2$, since we are in the ε -window. Combined with (37), for δ sufficiently small, we get that the total influence, for p in the ε -window, satisfies

$$I^p(f_n) = \mathbb{E}[|\text{Spec}^p(f_n)|] \geq c(\varepsilon)n^{1/3}, \quad c(\varepsilon) > 0,$$

giving that the width of the ε -window is at most $C(\varepsilon)n^{-1/3}$.

Finally, one might prefer to deal with transitive Boolean functions only, hence would want to consider high-dimensional percolation on tori. Another observation of [1] was that finite size boundary effects become important here, and the cluster structure in a torus is different from the cluster structure in a box. In particular, for percolation on the torus, already Erdős-Rényi random graph asymptotics take place: the largest critical cluster has size of order $n^{2d/3}$, and the critical window should be $n^{-d/3}$. (Nevertheless, large clusters are still “four-dimensional”, similarly to the box case; in particular, their Euclidean diameter, when pulled back to the universal cover of the torus, is $n^{d/6}$.) A large part of this conjectured basic near-critical picture has already been proved; see [5,6,24,25,30].

The above-mentioned critical window $n^{-d/3}$ should hold for most natural monotone events on the torus: e.g., for the existence of a non-contractible cluster, or a cluster of pulled-back Euclidean diameter $n^{d/6}$, or a cluster of size $n^{2d/3}$. However, the near-critical tails should be different. At the lower end of the window, with very negative λ , to get a cluster with large Euclidean diameter $n^{d/6}$ (independent of λ), the near-critical GW tree diameter asymptotics (33) should be relevant, yielding a tail $\exp(-\Theta(|\lambda|))$. To get a cluster with volume $n^{2d/3}$, the near-critical GW tree volume asymptotics should be relevant, which can be shown to be $\exp(-\Theta(|\lambda|^2))$. A similar difference for the Erdős-Rényi random graphs was pointed out in [40]. Regarding the tail for the existence of a non-contractible cluster, we will not make a guess. But in any case, it seems unlikely that natural monotone events will have subexponential tails.

For a recent survey of high-dimensional percolation and random graphs, see [26]. For a renormalization group physics approach, addressing finite size scaling both for free and periodic boundary conditions, see [33] and the references therein.

Acknowledgements

This work was begun while DA was visiting the Department for Mathematical Sciences, Gothenburg, to which he is grateful for their hospitality. The main authors thank Anders Martinsson for jointly proving part (b) of Theorem 3 and for other useful comments. They further thank Bo Berndtsson, Christophe Garban and Benjy Weiss for some discussions, and an anonymous referee for a number of comments on the manuscript.

The appendix author is grateful to Balázs Ráth for an eye-opening discussion and comments on the manuscript, and to the referee for pointing out that the proof of (29) can basically be found already in [43]. GP is also grateful to Christophe Garban and Alan Hammond for many discussions on near-critical and dynamical percolation, and to Remco van der Hofstad, Tim Hulshof, and Gady Kozma for conversations about the high-dimensional case.

DA was during this work supported by postdoctoral grants 150804/2012-1 from the Brazilian CNPq and 637-2013-7302 from the Swedish Research Council. JES also acknowledges the support of the Swedish Research Council and the Knut and Alice Wallenberg Foundation. GP was partially supported by the Hungarian National Research, Development and Innovation Office, NKFIH, grant K109684 and by the MTA Rényi Institute “Lendület” Limits of Structures Research Group.

References

- [1] M. Aizenman. On the number of incipient spanning clusters. *Nuclear Phys. B* **485** (3) (1997) 551–582. [MR1431856](#)
- [2] I. Benjamini, G. Kalai and O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Publ. Math. Inst. Hautes Études Sci.* **90** (1999) 5–43. [MR1813223](#)
- [3] B. Bollobás. *Random Graphs*, 2nd edition. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge University Press, Cambridge, 2001. [MR1864966](#)
- [4] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica* **7** (1) (1987) 35–38. [MR0905149](#)
- [5] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade and J. Spencer. Random subgraphs of finite graphs. I. The scaling window under the triangle condition. *Random Structures Algorithms* **27** (2) (2005) 137–184. [MR2155704](#)
- [6] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade and J. Spencer. Random subgraphs of finite graphs. II. The lace expansion and the triangle condition. *Ann. Probab.* **33** (5) (2005) 1886–1944. [MR2165583](#)
- [7] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Geom. Funct. Anal.* **7** (3) (1997) 438–461. [MR1466334](#)
- [8] H. Dumitriu-Copin. Limit of the Wulff crystal when approaching critically for site percolation on the triangular lattice. *Electron. Commun. Probab.* **18** (93) (2013) 9. [MR3151749](#)
- [9] R. Durrett. *Probability: Theory and Examples*, 4th edition. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010. [MR2722836](#)
- [10] E. M. Elçi. Personal communication, 2016.
- [11] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960) 17–61. [MR0125031](#)
- [12] R. Fitzner and R. van der Hofstad. Nearest-neighbor percolation function is continuous for $d > 10$. Preprint. Available at [arXiv:1506.07977](#).
- [13] E. Friedgut. Sharp thresholds of graph properties, and the k -sat problem. *J. Amer. Math. Soc.* **12** (4) (1999) 1017–1054. With an appendix by Jean Bourgain. [MR1678031](#)
- [14] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** (10) (1996) 2993–3002. [MR1371123](#)
- [15] C. Garban, G. Pete and O. Schramm. Pivotal, cluster, and interface measures for critical planar percolation. *J. Amer. Math. Soc.* **26** (4) (2013) 939–1024. [MR3073882](#)
- [16] C. Garban, G. Pete and O. Schramm. The scaling limits of near-critical and dynamical percolation. Preprint. Available at [arXiv:1305.5526](#).
- [17] C. Garban and J. E. Steif. Noise sensitivity and percolation. In *Probability and Statistical Physics in Two and More Dimensions* 49–154. *Clay Math. Proc.* **15**. Amer. Math. Soc., Providence, RI, 2012. [MR3025390](#)
- [18] C. Garban and J. E. Steif. *Noise Sensitivity of Boolean Functions and Percolation*. Cambridge University Press, Cambridge, 2015. [MR3468568](#)
- [19] G. Grimmett. *Percolation*, 2nd edition. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **321**. Springer-Verlag, Berlin, 1999. [MR1707339](#)
- [20] A. Hammond, E. Mossel and G. Pete. Exit time tails from pairwise decorrelation in hidden Markov chains, with applications to dynamical percolation. *Electron. J. Probab.* **17** (68) (2012) 16. [MR2968675](#)
- [21] T. Hara. Mean-field critical behaviour for correlation length for percolation in high dimensions. *Probab. Theory Related Fields* **86** (3) (1990) 337–385. [MR1069285](#)
- [22] T. Hara and G. Slade. Mean-field critical behaviour for percolation in high dimensions. *Comm. Math. Phys.* **128** (1990) 333–391. [MR1043524](#)
- [23] T. Hara and G. Slade. The scaling limit of the incipient infinite cluster in high-dimensional percolation, II. Integrated super-Brownian excursion. *J. Math. Phys.* **41** (3) (2000) 1244–1293. [MR1757958](#)
- [24] M. Heydenreich and R. van der Hofstad. Random graph asymptotics on high-dimensional tori. *Comm. Math. Phys.* **270** (2) (2007) 335–358. [MR2276449](#)
- [25] M. Heydenreich and R. van der Hofstad. Random graph asymptotics on high-dimensional tori II: Volume, diameter and mixing time. *Probab. Theory Related Fields* **149** (3–4) (2011) 397–415. [MR2776620](#)
- [26] M. Heydenreich and R. van der Hofstad. Progress in high-dimensional percolation and random graphs. Lecture notes for the CRM-PIMS Summer School in Probability 2015. Available at <http://www.win.tue.nl/~rhofstad/survey-high-d-percolation.pdf>.
- [27] M. Heydenreich, R. van der Hofstad and T. Hulshof. Random walk on the high-dimensional IIC. *Comm. Math. Phys.* **329** (1) (2014) 57–115. [MR3206998](#)

- [28] M. Heydenreich, R. van der Hofstad, T. Hulshof and G. Miermont. Backbone scaling limit of the high-dimensional IIC. In preparation.
- [29] R. van der Hofstad and A. A. Járai. The incipient infinite cluster for high-dimensional unoriented percolation. *J. Stat. Phys.* **114** (3–4) (2004) 625–663. [MR2035627](#)
- [30] R. van der Hofstad and A. Sapozhnikov. Cycle structure of percolation on high-dimensional tori. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (3) (2014) 999–1027. [MR3224297](#)
- [31] J. Kahn, G. Kalai and N. Linial. The influence of variables on Boolean functions. In *29th Annual Symposium on Foundations of Computer Science* 68–80, 1988.
- [32] G. Kalai and S. Safra. Threshold phenomena and influence: Perspectives from mathematics, computer science, and economics. In *Computational Complexity and Statistical Physics* 25–60. *St. Fe Inst. Stud. Sci. Complex.* Oxford Univ. Press, New York, 2006. [MR2208732](#)
- [33] R. Kenna and B. Berche. Universal finite-size scaling for percolation theory in high dimensions. Preprint. Available at [arXiv:1606.00315](#).
- [34] H. Kesten. Scaling relations for 2D-percolation. *Comm. Math. Phys.* **109** (1) (1987) 109–156. [MR0879034](#)
- [35] G. Kozma. Personal communication, 2016.
- [36] G. Kozma and A. Nachmias. The Alexander-Orbach conjecture holds in high dimensions. *Invent. Math.* **178** (2009) 635–654. [MR2551766](#)
- [37] G. Kozma and A. Nachmias. Arm exponents in high dimensional percolation. *J. Amer. Math. Soc.* **24** (2) (2011) 375–409. [MR2748397](#)
- [38] R. P. Langlands. The renormalization fixed point as a mathematical object. In *Twenty Years of Białowieża: A Mathematical Anthology* 185–216. *World Sci. Monogr. Ser. Math.* **8**. World Sci. Publ., Hackensack, NJ, 2005. [MR2181552](#)
- [39] R. P. Langlands and M.-A. Lafortune. Finite models for percolation. In *Representation Theory and Analysis on Homogeneous Spaces (New Brunswick, NJ, 1993)* 227–246. *Contemp. Math.* **177**. Amer. Math. Soc., Providence, RI, 1994. [MR1303608](#)
- [40] T. Łuczak. Random trees and random graphs. *Random Structures Algorithms* **13** (1998) 485–500. In Proceedings of the Eighth International Conference “Random Structures and Algorithms” (Poznan, 1997). [MR1662797](#)
- [41] E. Mossel and R. O’Donnell. On the noise sensitivity of monotone functions. *Random Structures Algorithms* **23** (3) (2003) 333–350. [MR1999039](#)
- [42] A. Nachmias and Y. Peres. The critical random graph, with martingales. *Israel J. Math.* **176** (1) (2010) 29–41. [MR2653185](#)
- [43] P. Nolin. Near-critical percolation in two dimensions. *Electron. J. Probab.* **13** (55) (2008) 1562–1623. [MR2438816](#)
- [44] R. O’Donnell, M. Saks, O. Schramm and R. Servedio. Every decision tree has an influential variable. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*, 31–39, 2005.
- [45] G. Pete. How long could it take to establish a crossing in dynamical percolation? Talk at MFO, Oberwolfach, September 2012. Available at <http://www.math.bme.hu/~gabor/DynExitOwolfach.pdf>.
- [46] R. Rossignol. Arbitrary threshold widths for monotone, symmetric properties. *J. Appl. Probab.* **44** (1) (2007) 164–180. [MR2312994](#)
- [47] L. Russo. An approximate zero-one law. *Z. Wahrsch. Verw. Gebiete* **61** (1) (1982) 129–139. [MR0671248](#)
- [48] O. Schramm and S. Smirnov. On the scaling limits of planar percolation. *Ann. Probab.* **39** (5) (2011) 1768–1814. With an appendix by Christophe Garban. [MR2884873](#)
- [49] O. Schramm and J. Steif. Quantitative noise sensitivity and exceptional times for percolation. *Ann. of Math. (2)* **171** (2) (2010) 619–672. [MR2630053](#)
- [50] G. Slade. Scaling limits and super-Brownian motion. *Notices Amer. Math. Soc.* **49** (9) (2002) 1056–1067. [MR1927455](#)
- [51] S. Smirnov and W. Werner. Critical exponents for two-dimensional percolation. *Math. Res. Lett.* **8** (5–6) (2001) 729–744. [MR1879816](#)
- [52] M. Talagrand. On Russo’s approximate zero-one law. *Ann. Probab.* **22** (3) (1994) 1576–1587. [MR1303654](#)
- [53] W. Werner. Lectures on two-dimensional critical percolation. In *Statistical Mechanics* 297–360. *IAS/Park City Math. Ser.* **16**. Amer. Math. Soc., Providence, RI, 2009. [MR2523462](#)