Information-theoretic proofs of zero-one laws

Authors: Yahya Ayach
Anthony Khairallah
Tia Manoukian
Jad Mchaimech
Adam Salha

Advisor: Siamak Taati

Final Report

Summer Research Camp 2023

Department of Mathematics and Center for Advanced Mathematical Sciences
American University of Beirut

Abstract

Neil O'Connell offered an elegant information-theoretic proof of the Hewitt-Savage zero-one law, which states that every exchangeable event concerning a sequence of independent and identically distributed random variables is deterministic (i.e., has a probability of either 0 or 1). We extend his argument to prove a new zero-one law for events that have a sufficient amount of symmetries. Our result encompasses the Hewitt-Savage law and the zero-one law for shift-invariant events, as well as examples that are not covered by earlier results. Furthermore, we relax the i.i.d. condition and allow for the random variables to be only asymptotically independent.

1 Introduction

1.1 Introduction to Zero-One Laws

Zero-one laws are a class of theorems in probability theory that assert that certain types of events are deterministic, i.e., the probability of their occurrence is 0 or 1. The events we are working with are concerned with an infinite sequence of random variables, $(X_1, X_2, ...)$.

Kolmogorov's zero-one law is concerned with tail events. A tail event is one such that its occurrence or non-occurrence is not affected by changing finitely many terms. The theorem states that given a sequence of independent random variables, any tail event concerning those random variables is deterministic. An interesting example of a tail event is percolation. Picture the lattice of points on the plane that have integer coordinates. Between every two adjacent points on the lattice, there is a probability p that an edge connecting them exists, where the probabilities of edges existing are independent. Percolation is the phenomenon where there exists a connected path of edges to infinity, i.e., an infinitely long path. This is a tail event since it is independent of changing any finite number of terms. An infinite path will remain infinite if finitely many edges are added or removed, and so does the finite one. So, the percolation event is deterministic.

The Hewitt-Savage zero-one law is concerned with exchangeable events. An exchangeable event is one whose occurrence is not affected by applying a permutation on a finite subset of terms. The Hewitt-Savage zero-one law states that exchangeable events containing sequences with independent identically distributed (i.i.d.) random variables. An interesting example of an event covered by this theorem is in the context of a random walk. Starting from the origin of the number line, at each step i, the walker either takes a step to the right ($X_i = 1$) with probability p or a step to the left ($X_i = -1$) with probability $1 - p$, with $p \in (0, 1)$. We are interested in the event that the walker returns to the origin infinitely many times, in other words, $\sum_{i=1}^{k} X_i = 0$ for infinitely many k's. It is clear that exchanging any finite number of X_i's will not affect the occurrence or non-occurrence of this event. Therefore, the probability of the walker returning to the origin infinitely many times is either 0 or 1.

The shift-invariance zero-one law (a.k.a. the ergodicity of an i.i.d. process) states that all shift-invariant events concerning a sequence of i.i.d. random variables — events whose occurrences are not affected by shifting the terms — are also deterministic. An example of a shift-invariant event is the one addressed in the infinite monkey theorem. A monkey is sitting behind a keyboard, typing characters from a finite set of characters one after another, completely at random, and independently of each other. According to the infinite monkey theorem, if we wait long enough, the monkey will almost surely (i.e., with probability 1) type Shakespeare's entire text of Hamlet, infinitely many times. The shift-invariance zero one law gives the weaker conclusion that the probability of this event is either 0 or 1. Indeed, we can see that shifting the random chain of keys hit would not affect the appearance of the text infinitely many times. Thus, this event is deterministic.

These three theorems were originally proved using measure theory. However, about twenty years ago, Neil O'Connell discovered an information-theoretic proof of the Hewitt-Savage theorem. O'Connell's proof is based on an information-theoretic inequality, which we call O'Connell's inequality. We will reproduce O'Connell's proof in Section 1.3 below after reviewing the basic concepts of information theory.

O'Connell's argument can be easily adapted to give a proof of the shift-invariance zero-one law. We will generalize his argument to prove a theorem that covers the latter two zero-one laws as well as some other interesting examples which were not covered by any of the above three theorems. Our theorem states that if an event concerning an infinite sequence of i.i.d. random variables has "enough symmetries," then it is deterministic. We then extend our argument by relaxing the i.i.d. condition to only require the random variables to be asymptotically independent and have close distributions. Finally, we started examining the case when the sequence of random variables is finite. We considered two examples of events that are exchangeable and studied their probabilities of occurrence.

1.2 Basic Information Theory

Since the argument relies on information theory, we will introduce some concepts which will be used in the proofs.

We begin by defining entropy. Informally, entropy measures the amount of missing information or uncertainty regarding the value of a random variable. It can roughly be thought of as the average number of yes/no questions needed to determine the value of a random variable, given its probability mass function.

Definition 1 (Entropy)**.** The *entropy* of a discrete random variable X with values in a finite set $\mathcal{X}$ and probability mass function p is defined as

$$H(X) = -\sum_{x\in\mathcal{X}} \mathbb{P}(x)\log\mathbb{P}(x)\ .$$

Throughout this report, all the logarithms are in base 2.

The conditional entropy of a random variable Y given another random variable X is the remaining uncertainty in Y if we know the value of X.

Definition 2 (Conditional entropy)**.** Given two random variables X and Y with values from the finite sets $\mathcal{X}$ and $\mathcal{Y}$ the *conditional entropy* of Y given X is defined as:

$$\begin{aligned} H(Y \mid X) &= \sum_{x\in\mathcal{X}} \mathbb{P}(X=x)H(Y \mid X=x) \\ &= -\sum_{x\in\mathcal{X}} p_X(x)\sum_{y\in\mathcal{Y}} \mathbb{P}(Y=y \mid X=x)\log\mathbb{P}(Y=y \mid X=x) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{P}(X=x, Y=y)\log\mathbb{P}(Y=y \mid X=x)\ . \end{aligned}$$

Observe that $H(X)$ is determined by the distribution of X, and $H(Y \mid X)$ is determined by the joint distribution of X and Y. We next define the relative entropy of two probability distributions. Informally speaking, relative entropy measures how much one distribution deviates from another distribution.

Definition 3 (Relative entropy)**.** The *relative entropy* between two probability mass functions $p(x)$ and $q(x)$ on a finite set $\mathcal{X}$ is defined as

$$D(p \parallel q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}\ .$$

We now define the mutual information between two random variables X and Y. It can be thought of as how much we learn about X from knowing the value of Y, or in other words, the amount of shared information between X and Y.

Definition 4 (Mutual information)**.** Given two random variables X and Y, the *mutual information* between X and Y is defined as

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

Further background on the properties of entropy, mutual information, and relative entropy can be found in the book by Cover and Thomas [4].

1.3 O'Connell's Proof of the Hewitt-Savage Zero-One Law

We first formally define what it means for an event to be *exchangeable*.

Definition 5 (Exchangeable event)**.** An event $E \in M^{\mathbb{N}}$ is *exchangeable* if for every finitary permutation $\pi : \mathbb{N} \to \mathbb{N}$, we have

$$(a_n)_{n\in\mathbb{N}} \in E \quad\Longleftrightarrow\quad (a_{\pi(n)})_{n\in\mathbb{N}} \in E\ .$$

Here, a *permutation* of $\mathbb{N}$ refers to a bijective map $\mathbb{N} \to \mathbb{N}$. A permutation $\pi : \mathbb{N} \to \mathbb{N}$ is *finitary*, if there is a finite set $J \subseteq \mathbb{N}$ such that $\pi(k) = k$ for $k \in \mathbb{N} \setminus J$.

Theorem 1.1 (Hewitt-Savage Zero-One Law)**.** *Let $\underline{X} = (X_1, X_2, \ldots)$ be a sequence of i.i.d. random variables with values in a set M. Then, every exchangeable event $E \subseteq M^{\mathbb{N}}$ is deterministic, that is, $\mathbb{P}(\underline{X} \in E) \in \{0, 1\}$.*

O'Connell's proof involves only one result from measure theory, formulated in the following lemma.

Lemma 1.2. *Let $Y, X_1, X_2, \ldots$ be random variables, and suppose that for every n, the random variable Y is independent of $(X_1, X_2, \ldots, X_n)$. Then, Y is independent of the entire sequence $(X_1, X_2, \ldots)$.*

The proof of the theorem mainly relies on the following information-theoretic inequality.

Lemma 1.3 (O'Connell's Inequality)**.** *Let $V, W_1, W_2, \ldots, W_n$ be discrete random variables and suppose that $W_1, W_2, \ldots, W_n$ are independent. Then,*

$$H(V) \geq \sum_{i=1}^{n} I(V; W_i)$$

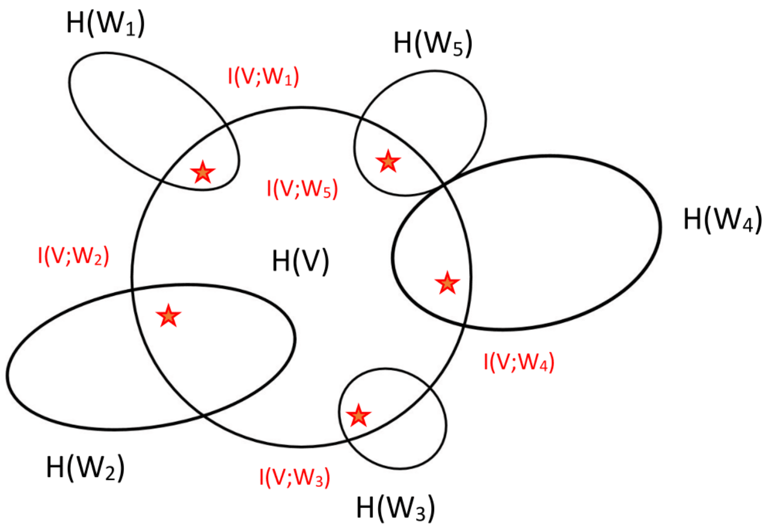

Figure 1: Visual demonstration of O'Connell's Inequality. Here, the regions represent the information needed to know each event (e.g., H(V)), and the shared regions are the mutual information that we gain about V when we know W_n. As you can see, there are no shared regions between any W_i and W_j, $j \neq i$ since they are independent and have no mutual information. The figure demonstrates that what we know about each or all W_n could at most contribute to knowing all of V, but not necessarily.

Proof. Figure 1 provides a visual proof of the lemma. We prove it rigorously as follows:

$$\begin{aligned}
H(V) &\geq H(V) - H(V \mid (W_1, W_2, \ldots, W_n)) \\
&= I\big(V; (W_1, W_2, \ldots, W_n)\big) \\
&= H(W_1, W_2, \ldots, W_n) - H(W_1, W_2, \ldots, W_n \mid V) \\
&= H(W_1) + H(W_2 \mid W_1) + H(W_3 \mid W_1, W_2) + \cdots + H(W_n \mid W_1, W_2, \ldots, W_{n-1}) \\
&\qquad - H(W_1 \mid V) - H(W_2 \mid W_1, V) - \cdots - H(W_n \mid W_1, W_2, \ldots, W_{n-1}, V) \\
&\geq H(W_1) + H(W_2) + \cdots + H(W_n) - H(W_1 \mid V) - H(W_2 \mid V) - \cdots - H(W_n \mid V) \\
&= \big[H(W_1) - H(W_1 \mid V)\big] + \big[H(W_2) - H(W_2 \mid V)\big] + \cdots + \big[H(W_n) - H(W_n \mid V)\big] \\
&= I(V; W_1) + I(V; W_2) + \cdots + I(V; W_n)\ .
\end{aligned}$$

The equality on the fourth line is by the chain rule. For the inequality on the fifth line, we have used the independence of $W_1, W_2, \ldots, W_n$ and the fact that conditioning can only reduce the entropy. □

Proof of Theorem 1.1. This proof is due to O'Connell [7]. First, we apply Lemma 1.3 with $V := \mathbb{1}_E(\underline{X})$ and $W_i = X_i$ for $i = 1, 2, \ldots, n$ and arbitrary n:

$$H(\mathbb{1}_E(\underline{X})) \geq I(\mathbb{1}_E(\underline{X}); X_1) + I(\mathbb{1}_E(\underline{X}); X_2) + \cdots + I(\mathbb{1}_E(\underline{X}); X_n)$$

We claim that $I(\mathbb{1}_E(\underline{X}); X_i) = I(\mathbb{1}_E(\underline{X}), X_1)$ for all i.

Since the mutual information between two random variables U and V only depends on their joint distribution, it suffices to show that the joint distribution of $\mathbb{1}_E(\underline{X})$ and X_i is the same for all i. To show the latter, note that for $a \in \{0, 1\}$ and $b \in M$,

$$\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_i = b) = \mathbb{P}(\mathbb{1}_E(X_i, X_2, \ldots, X_1, X_{i+1}, \ldots) = a, X_1 = b)$$

since $\{X_i\}$ are i.i.d. The latter is equal to $\mathbb{P}(\mathbb{1}_E(\underline{X})) = a, X_1 = b)$ since E exchangeable.

We conclude that,

$$H(\mathbb{1}_E(\underline{X})) \geq nI(\mathbb{1}_E(\underline{X}); X_1)$$

Since $\mathbb{1}_E(\underline{X})$ is a Bernoulli random variable, $H(\mathbb{1}_E(\underline{X})) \leq 1$. Hence, $I(\mathbb{1}_E(\underline{X}); X_1) \leq \frac{1}{n} \to 0$ as $n \to \infty$. It follows that $\mathbb{1}_E(\underline{X})$ and X_1 are independent, implying that E and X_1 are independent.

Next, we apply Lemma 1.3 with $V = \mathbb{1}_E$ and $W_i = (X_{2i-1}, X_{2i})$ for $i = 1, 2, \ldots, n$ and arbitrary n, and we can similarly show that E and (X_1, X_2) are independent.

Similarly, we do this for the n-tuple $(X_1, X_2, \ldots, X_n)$ of arbitrary size n add to show that E is independent of $(X_1, X_2, \ldots, X_n)$ for every n. Using Lemma 1.2, we get that E and $(\underline{X})\underline{X}$ are independent, meaning that E is independent of any possible event concerning $\underline{X}$; in particular, E is independent of itself. Hence,

$$\mathbb{P}((X_n)_{n\in\mathbb{N}} \in E \cap E) = \mathbb{P}((X_n)_{n\in\mathbb{N}} \in E)\mathbb{P}((X_n)_{n\in\mathbb{N}} \in E) ,$$

which implies $\mathbb{P}((X_n)_{n\in\mathbb{N}} \in E)$ must be either 0 or 1. □

2 Zero-one laws for sufficiently symmetric events regarding i.i.d. collections

2.1 Main Theorem

Instead of limiting our scope to specific symmetries such as exchangeability and shift invariance, we will now look at a more general class of symmetries, which we call positional symmetries.

Definition 6. Let Γ be a countable set. A *positional symmetry* of an event $E \subseteq M^\Gamma$ is a map $\pi : \Gamma \to \Gamma$ such that

$$(a_n)_{n\in\Gamma} \in E \quad \Longleftrightarrow \quad (a_{\pi(n)})_{n\in\Gamma} \in E \; .$$

We remark that we do not require the map π to be bijective.

We now generalize the Hewitt-Savage zero-one law and the zero-one law for shift-invariant events to include a larger set of events. The following theorem states that events that have enough positional symmetries are deterministic.

Theorem 2.1. *Let $\underline{X} := (X_n)_{n\in\Gamma}$ be a family of random variables with values from a finite set M. Consider an event $E \subseteq M^\Gamma$, and suppose that for every finite $J \subseteq \Gamma$, the set E has an injective positional symmetry $\pi : \Gamma \to \Gamma$ such that $\pi(J) \cap J = \varnothing$. Then E is deterministic; that is, $\mathbb{P}\big(\underline{X} \in E\big) \in \{0, 1\}$.*

Proof. The proof follows the same argument as that of Theorem 1.1. The only thing worth noting is the proof that for each size of tuples of random variables, we have infinitely many disjoint tuples all of which have the same joint distribution with $\mathbb{1}_E(\underline{X})$.

Namely, let $I \subseteq \Gamma$ be a finite set and consider the tuple $X_I := (X_i)_{i\in I}$. We denote the indicator of E be g. Let us show that there exists an infinite sequence $I_0, I_1, I_2, \dots$ of disjoint finite subsets $I_k \in \Gamma$ with $I_0 := I$ and $|I_k| = |I|$ such that the joint distributions of the pairs $(g(\underline{X}), X_{I_k})$ are all the same.

We start with an injective positional symmetry π_1 corresponding to $J = I_0$. The reasoning used in the proof of Theorem 1.1 shows that

$$(\mathbb{1}_E, X_I) \sim (g(\underline{X}), X_{I_0}) \sim (g(\underline{X}), X_{\pi_1(I_0)})$$

Let $I_1 := \pi_1(I_0)$. Clearly, I_1 and I_0 are disjoint.

Next, let π_2 be an injective positional symmetry corresponding to $J = I_0 \cup I_1$. Then, the same reasoning gives

$$(\mathbb{1}_E, X_I) \sim (g(\underline{X}), X_{I_0}) \sim (g(\underline{X}), X_{\pi_2(I_0)})$$

Let $I_2 := \pi_2(I_0)$. The assumption $\pi_2(J) \cap J = \varnothing$ ensures that I_2 is disjoint from I_0 and I_1.

By repeating the above process, letting π_n to be an injective positional symmetry corresponding to $J = \bigcup_{k=1}^{n-1} I_k$ and setting $I_n := \pi_n(I_0)$, we obtain an finite sequence $I_0, I_1, I_2, \dots$ with the desired properties.

Since the distribution of $(g(\underline{X}), X_{I_k})$ is the same for all k, we find that the mutual information between $g(\underline{X})$ and X_{I_k} is also the same for all k. Thus, Lemma 1.3 implies that $g(\underline{X})$ is independent of X_I. Since I is arbitrary, Lemma 1.2 implies that $g(\underline{X})$ is independent of $\underline{X}$, hence also independent of itself. It follows, as before, that $\mathbb{P}(\underline{X} \in E) \in \{0, 1\}$.

□

2.2 Consequences of the Theorem

2.2.1 Mentioned Zero-One Laws

The first corollary of the theorem is the Hewitt-Savage zero-one law.

Alternative proof of Theorem 1.1. Let $E \subseteq M^\Gamma$ be an exchangeable event. Then for every finite $J \subseteq \Gamma$, we can choose a finitary permutation $\pi : \Gamma \to \Gamma$ such that $\pi(J) \cap J = \varnothing$. Furthermore, since E is exchangeable, we have

$$(X_i)_{i\in\Gamma} \in E \quad \Longleftrightarrow \quad (X_{\pi(i)})_{i\in\Gamma} \in E \; ,$$

hence π is a positional symmetry of E. Thus, Theorem 2.1 implies that E is deterministic. □

Another result is the zero-one law for shift-invariant events. We start by defining shift-invariance.

Definition 7. An event $E \in M^{\mathbb{N}}$ is *shift-invariant* if for every $k \in \mathbb{N}$,

$$(a_n)_{n\in\mathbb{N}} \in E \iff (a_{k+n})_{n\in\mathbb{N}} \in E \ .$$

Again, we can show that the shift-invariant zero-one law follows from our theorem.

Corollary 2.1.1 (Shift-invariance zero-one law)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with values in a set M. Then, every shift-invariant event $E \subseteq M^{\mathbb{N}}$ is deterministic, that is, $\mathbb{P}\big((X_n)_{n\in\mathbb{N}} \in E\big) \in \{0,1\}$.*

Proof. Let $E \subseteq M^{\Gamma}$ be a shift-invariant event. Given a finite set $J \subseteq \Gamma$, we let k be the largest element of J and define $\pi : \Gamma \to \Gamma$ by $\pi(n) = n + k$. Then, π is injective and satisfies $J \cap \pi(J) = \varnothing$. Furthermore, by the shift-invariance of E, we have

$$(X_i)_{i\in\Gamma} \in E \iff (X_{\pi(i)})_{i\in\Gamma} \in E \ ,$$

hence π is a positional symmetry of E. Therefore, Theorem 2.1 implies that E is deterministic. $\square$

2.2.2 Graph-Theoretic Zero-One Laws

The following corollaries are concerned with graph theory. We first need the following two definitions.

Definition 8. An infinite Erdős-Rényi Random Graph with parameter $p \in [0,1]$ is a graph with a countably infinite vertex set V (for simplicity, we assume $V = \mathbb{N}$) such that between every two distinct vertices, there is an edge with probability p, independently of the other pairs.

Such a graph can be represented by an array of i.i.d. Bernoulli random variables $(X_{i,j})_{i,j\in\mathbb{N},i<j}$ with parameter p, where $X_{i,j}$ indicates whether the two vertices i and j are connected by an edge or not.

Definition 9. Two graphs $G = (V,E)$ and $H = (W,F)$ are said to be *isomorphic* if there exists a bijective function $h : V \to W$ such that $ab \in E \iff h(a)h(b) \in E$.

A *graph-theoretic property* P is a property where if two graphs G and H are isomorphic, then $G \in P \iff H \in P$.

For example, connectivity is a graph-theoretic property. Note that, as a property of the graphs on the vertex set $V = \mathbb{N}$, connectivity is an event concerning $(X_{i,j})$. It consists of all connected graphs.

The following result is well-known and already has a proof which we will discuss after relating it to positional symmetries.

Corollary 2.1.2 (Graph-theoretic zero-one law)**.** *Every graph-theoretic property of the countably-infinite Erdős-Rényi is deterministic.*

Proof. Here, the event E is the graph-theoretic property P and Γ is $\mathbb{N} \times \mathbb{N}$. Let J be a finite subset of $\mathbb{N} \times \mathbb{N}$ and $\widetilde{J}$ be the set of vertices included in J. Let Φ be a permutation on the elements of $\widetilde{J}$ such that $\Phi(\widetilde{J}) \cap \widetilde{J} = \varnothing$. Our injective positional symmetry will be $\pi : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, where $\pi(i,j) = (\Phi(i), \Phi(j))$. Since $\Phi(\widetilde{J}) \cap \widetilde{J} = \varnothing$, we get $\pi(J) \cap J = \varnothing$ $\square$

The standard proof relies on the following lemma.

Lemma 2.2. *Let $G = (X_{i,j})_{i,j\in\mathbb{N},i<j}$ and $H = (Y_{i,j})_{i,j\in\mathbb{N},i<j}$ be two independent countably-infinite Erdős-Rényi Random Graphs. Then $\mathbb{P}(G$ and H are isomorphic$) = 1$.*

Alternative proof of Corollary 2.1.2. Let P be a graph-theoretic property. Since the Bernoulli random variables are i.i.d.,

$$\mathbb{P}((X_{i,j}) \in P \text{ and } (Y_{i,j}) \in P) = \mathbb{P}((X_{i,j}) \in P)^2 \ .$$

Moreover, by Lemma 2.2,

$$\mathbb{P}((X_{i,j}) \in P \text{ and } (Y_{i,j}) \in P) = \mathbb{P}((X_{i,j}) \in P) \ .$$

Thus, $\mathbb{P}((X_{i,j}) \in P)^2 = \mathbb{P}((X_{i,j}) \in P)$, and therefore, $\mathbb{P}((X_{i,j} \in P) \in \{0,1\}$. $\square$

The first argument for Corollary 2.1.2 can be repeated to obtain zero-one laws for Bernoulli random sub-graphs of any other infinite graph that has enough symmetries.

Given a graph $H = (V, E)$ and a value $p \in [0, 1]$, we can construct a Bernoulli random sub-graph of G by keeping each edge with probability p and removing it with probability $1-p$, independently of the other edges. In this framework, the countably-infinite Erdős-Rényi graph is a Bernoulli random sub-graph of the complete graph on vertex set $\mathbb{N}$.

An *automorphism* (or *symmetry*) of a graph $G = (V, E)$ is an isomorphism of G with itself, that is, a bijective map $\theta : V \to V$ such that $ab \in E$ if and only if $\theta(a)\theta(b) \in E$.

Corollary 2.2.1 (Zero-one law for graph-theoretic properties of sufficiently symmetric infinite graphs)**.** *Let $H = (\mathbb{N}, E)$ be a countably-infinite graph. Suppose that for every finite set $A \subseteq \mathbb{N}$ of vertices, H has an automorphism θ such that $\theta(A) \cap A = \varnothing$. Then, for every $p \in [0, 1]$, every graph-theoretic property of a random subgraph of H with parameter p is deterministic.*

Proof. Consider the family $(X_e)_{e\in E}$ of Bernoulli random variables, where X_e indicates whether the edge e is kept or removed. By definition, these random variables are i.i.d. with parameter p.

Let $J \subseteq E$ be a finite set. Let $A \subseteq V$ denote the set of vertices incident to J. Clearly, A is also a finite set. Let θ be an automorphism of H satisfying $\theta(A) \cap A = \varnothing$, and define $\pi : E \to E$ by $\pi(ab) := \theta(a)\theta(b)$ for every $ab \in E$. Note that $\theta(a)\theta(b) \in E$ if and only if $ab \in E$ because θ is an automorphism of H. Furthermore, π is injective because θ is. Lastly, $\pi(J) \cap J = \varnothing$ because the vertices incident to J and $\pi(J)$ are disjoint.

Now, consider a graph-theoretic property P. We can view P as a (measurable) subset of $\{0, 1\}^E$. Since θ is an automorphism of H, a subgraph G of H satisfies P if and only if its image $\theta(G)$ satisfies P. It follows that for every $(x_e)_{e\in E} \in \{0, 1\}^E$, we have $(x_e)_{e\in E} \in P$ if and only if $(x_e)_{\pi(e)\in E} \in P$, that is, π is an injective positional symmetry of P. It follows from Theorem 2.1 that the subgraph represented by $(X_e)_{e\in E}$ satisfies P either with probability 0 or with probability 1. □

Here are some examples of infinite graphs that satisfy the hypothesis of Corollary 2.2.1:

(i) The complete graph on $\mathbb{N}$. The random subgraph is simply the infinite Erdős-Rényi random graph. See [2, 3] for the discussion of various interesting properties of this random graph.

(ii) The complete bipartite graph on $\mathbb{N} \times \mathbb{N}$.

(iii) The hypercubic lattice of any dimension (i.e., $\mathbb{Z}^d$ with nearest-neighbour edges, where $d \in \mathbb{N}$). In this case, all graph-theoretic properties are shift-invariant, hence the zero-one law follows from the ergodicity of bond percolation on $\mathbb{Z}^d$ (i.e., the Bernoulli process on its edges). More generally, any Cayley graph of any infinite finitely-generated graph has the same property.

(iv) The universal triangle-free graph H_3. This is the unique (up to isomorphism) graph on a countable vertex set that does not contain any triangle (i.e., a complete subgraph on 3 vertices) but contains all finite triangle-free graphs as induced subgraphs [6]. Observe that a random subgraph of H_3 (for $0 < p \leq 1$) is still triangle-free and contains (with probability 1) every finite triangle-free graph as an induced subgraph, hence it is isomorphic to H_3 itself. The zero-one law for the graph-theoretic properties of such a random subgraph thus follows immediately from its uniqueness.

2.2.3 A New Class of Deterministic Events

We have proved in the last subsection that the classes of deterministic events that are exchangeable and shift-invariant have the positional symmetry property. Are there classes of deterministic events that have the positional symmetry property but are not exchangeable, shift-invariant, nor tail events? Fortunately, there are.

Let $\underline{X} = (X_i)_{i\in\mathbb{N}}$ be a sequence of independent Bernoulli random variables with the same parameter p. Consider an event of the form

$$E = \{(a_n) : \text{for all } m \text{ and } k, \text{ the sequence } (a_{(k-1)2^m+1}, ..., a_{k2^m}) \text{ has at least } h(m) \text{ ones}\} \subseteq \{0, 1\}^{\mathbb{N}} ,$$

where $h : \mathbb{N} \to \mathbb{N}_0$ and $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$.

Observe that $\mathbb{P}(\underline{X} \in E) = 0$ by the infinite monkey principle. We now argue that the fact that $\mathbb{P}(\underline{X} \in E) \in \{0, 1\}$ follows from Theorem 2.1. On the other hand, as long as h is not the constant 0 and $h(m) < 2^m$ for every m, the event E is not exchangeable, shift-invariant, or a tail event.

Proposition 2.1. *E has the positional symmetry property.*

Proof. For every finite $J \subseteq \Gamma$, we have an injective positional symmetry $\pi : \Gamma \to \Gamma$ characterized in the following way: We choose $k = 1$ (the first block) and m large enough such that all the indices in J are contained within $1, \ldots, 2^m$) and then let π be the finitary permutation that switches the whole sequence of these indices with that of $(2^m + 1, \ldots, 2^{m+1})$ and leaves everything else intact. Thus, we have the new indices of the random variables all not in J since they were all in the sequence prior. Furthermore, the sequence still belongs to E since every lower-order block within $(a_1, \ldots, a_{2^m})$ and $(a_{2^m+1}, ..., a_{2^{m+1}})$ still satisfies h (because only the positions were changed and every sub-sequence is intact but only moved) and the sequences themselves still do, as well as the sequence $(a_1, ..., a_{2^{m+1}})$ that contains both anyways. Furthermore, if the switched sequence belongs to E then we can simply switch it back so that the original one belongs to E by the same argument.

Hence, E has the positional symmetry property. □

Now, since we know that E is deterministic under having the positional symmetry property, we want to find conditions on E that make it not exchangeable, shift-invariant, nor a tail event. Note that if $h(m) = 0$ for every m, then E is the entire set $\{0, 1\}^{\mathbb{N}}$. On the other hand, if $h(m) \geq 2^m$ for some m, then E is a singleton, containing only the all-1 sequence. In either of these cases, E is a tail event, exchangeable and shift-invariant. So, let us assume that $h(m) > 0$ for some m, and $h(m) < 2^m$ for all m. In this case, E is clearly not a tail event since it is dependant on the positions of the zeros in the sequence. (For instance, if we start with a sequence in E, take m with $h(m) > 0$, and fill a 2^m-block with zeros, then we get a sequence outside E.) It is also not shift-invariant because, if m is such that $h(m) > 0$, then every element of E is a shift (by 2^m) of a sequence that is not in E, namely the one obtained by adding 2^m zeros at the beginning. It remains to verify that E is not exchangeable.

Define $\theta \coloneqq \min\{2^m - h(m) : \forall m, h(m) \neq 0\}$.

Proposition 2.2. *E is exchangeable if and only if every sequence $(a_n) \in E$ has at most θ zeros.*

Proof. We prove it in two parts:

- Part 1: [⇐]

  If any sequence in E has at most θ zeros, then for every m such that $h(m) \neq 0$, we have $\theta \leq 2^m - h(m)$, so $h(m) \leq 2^m - \theta$. Thus, if we permute the zeros in the sequence to any position, in the worst case to the same block $((k-1)2^m + 1, \ldots, k2^m)$, we would still have at least $h(m)$ ones.

- Part 2: [⇒]

  We prove the contra-positive. Namely, suppose the number of zeros in some sequence $(a_n) \in E$ is more than θ. We show that E is not exchangeable.

  Denote $\omega(\underline{a})$ as the number of zeros in $\underline{a} = (a_n)$. If $\forall \underline{a} \in E$, $\omega(\underline{a}) > \theta$, then $\omega(\underline{a}) > \min\{2^m - h(m) : h(m) \neq 0\}$. Suppose this minimum is attained at m_0. Hence, for some $\underline{a} \in E$, $\omega(\underline{a}) > 2^{m_0} - h(m_0)$.

  However, to have E exchangeable, we need to be able to permute these zeros to any position, including the worst case scenario as before, for m_0. Then, we must have $\omega(\underline{a}) + h(m_0) \leq 2^{m_0}$ for some $\underline{a} \in E$, i.e. $\omega(\underline{a}) \leq \theta$ for some $\underline{a} \in E$.

  But the opposite is true, $\omega(\underline{a}) > \theta, \forall \underline{a} \in E$, so E is not exchangeable. □

Hence, if we choose h such that $h(m) > 0$ for some m, $h(m) < 2^m$ for all m, and $\max_{\underline{a} \in E} \omega(\underline{a}) > \theta$, then E is not exchangeable, shift-invariant, nor a tail event.

Therefore, we have discovered a new class of deterministic events whose determinism follows from our theorem but not from the three zero-one laws for tail events, exchangeable events, and shift-invariant events.

2.2.4 Zero-One Law for Renormalization Transformations

We will now look at another example of a class of events that are seen to be deterministic by our theorem, but not by the previously stated zero-one laws.

Consider a sequence $\underline{X}^0 = (X_n^0)_{n \in \mathbb{Z}}$ where the X_i^0's are i.i.d. Bernoulli random variables with parameter p. Let $\underline{X}^1 = (X_n^1)_{n \in \mathbb{Z}}$ where

$$X_i^1 = \mathrm{maj}(X_{3i-1}^0, X_{3i}^0, X_{3i+1}^0)$$

In other words, the sequence consists of the majority elements of each block of size 3, starting form the block centered at 0. We repeat this process infinitely many times to get $\underline{X}^2$, $\underline{X}^3$, ... We are interested in the event E that the value of the random variable at position 0 eventually stabilizes at $X_0 = 1$ or 0. Define B_i^n to be the sub-sequence of elements of $\underline{X}^0$ which are transformed into X_i^n in the sequence $\underline{X}^n$.

For every finite set $J \subseteq \mathbb{Z}$, we take the smallest such block containing all the indices in J; the smallest n so that $J \subseteq B_i^n$ for some i. B_i^n is a sub-block of $B_j^{n+1} = B_A$ for some j, along with two other sub-blocks, call them B_B and B_C.

We now permute the three blocks B_A, B_B, B_C by a map $\pi : \mathbb{Z} \to \mathbb{Z}$. Notice that $\pi(J) \cap J = \varnothing$ because all the elements of J, which are in B_A, are now in a different block B_B or B_C.

Moreover, the map π is an injective positional symmetry of the event E, because permuting the blocks B_A, B_B, B_C will only permute the values of three X_i^n for three consecutive i's, preserving the majority element of the three, X_j^{n+1}. Thus, the occurrence of the event that X_0 eventually stabilizes is not affected because the sequence is not affected after the $(n+1)$st step, making π a positional symmetry.

Therefore, we have found a positional symmetry π for every finite $J \subseteq \mathbb{Z}$ so that J and $\pi(J)$ are disjoint. We conclude by Theorem 2.1 that the probability that the central element stabilizes eventually is either 0 or 1.

In fact, we can precisely calculate when the sequence stablizes to 0, 1 or neither. We start by looking at $\mathbb{P}(X_0^1 = 1)$ which we can easily calculate as $3p^2(1-p) + p^3 = p^2(3-2p)$. We will call this $f(p)$.

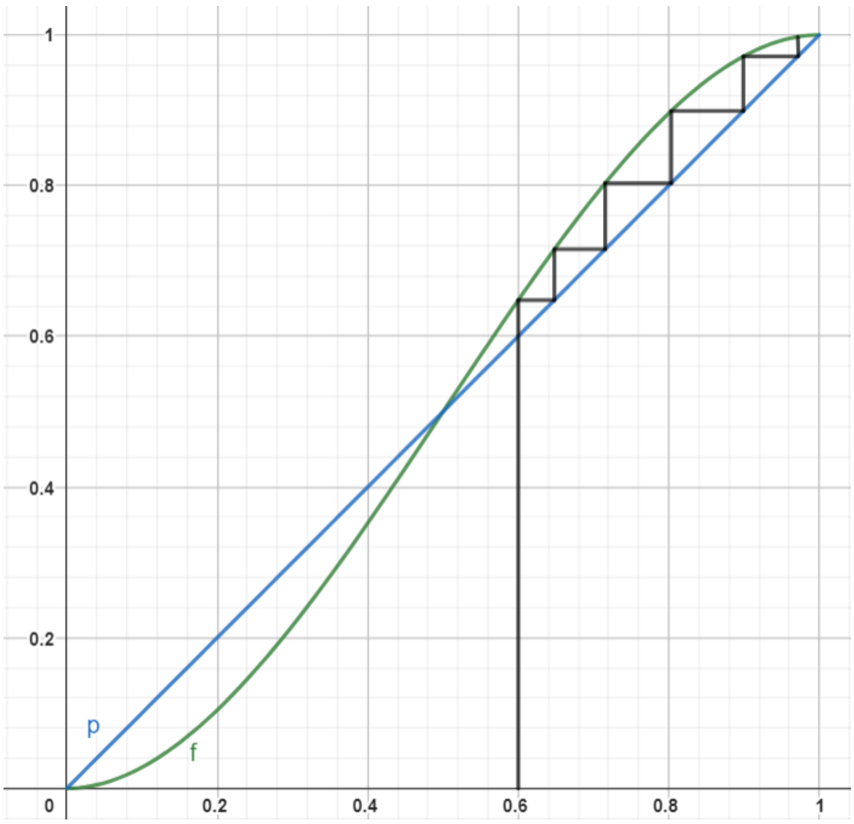

Figure 2: We take a $p > 0.5$. At each step, we consider $f(p)$ to be our new p, and keep applying f...

We notice that by applying f on p several times, the result will converge towards 1 when $p > \frac{1}{2}$ and 0 when $p < \frac{1}{2}$, and it will remain $\frac{1}{2}$ when $p = \frac{1}{2}$. So when $p \neq \frac{1}{2}$, the probability of stabilizing is 1, and when $p = \frac{1}{2}$, it is 0.

Note that the choice of taking the majority of 3 elements was arbitrary, and we could have chosen any size for the blocks. Also, we could have chosen a function different than the majority, as long as it does not depend on on the order of its parameters. However, it becomes harder to calculate exactly when the probability is 0 or 1.

3 Relaxing the i.i.d. Condition

In Theorem 2.1, we exploited the following two conditions on the random variables:

(i) X_J and $X_{\pi(J)}$ are independent,

(ii) The families $(X_n)_{n\in\Gamma}$ and $(X_{\pi(n)})_{n\in\Gamma}$ have the same distribution.

The former was ensured by the independence of X_is and the assumption $J \cap \pi(J) = \varnothing$; the latter was guaranteed by the injectivity of π and the i.i.d. assumption. We will relax condition (i) in Section 3.1 then condition (ii) in Section 3.2.

3.1 Relaxing the Condition of Independence

In this section, we will generalize the main theorem by considering random variables which are not necessarily independent, but rather "almost independent," or asymptotically independent.

To do so, we introduce the idea of ε-dependence. We can think of ε as a degree of dependence, so for instance 0-dependence is equivalent to independence, and the larger ε is the more dependence we have.

Definition 10. Let $\varepsilon > 0$.

- Two events E and E' are *ε-dependent* if $|\mathbb{P}(E \cap E') - \mathbb{P}(E)\mathbb{P}(E')| \leq \varepsilon$
- Two random variables X and Y are *ε-dependent* if for every measurable sets A and B, the events $X \in A$ and $Y \in B$ are ε-dependent, that is,

$$|\mathbb{P}(X \in A \text{ and } Y \in B) - \mathbb{P}(X \in A)\mathbb{P}(Y \in B)| \leq \varepsilon$$

We can also think of ε-dependence in terms of the mutual information. We will now show that the second interpretation follows from the definition. The two interpretations are actually equivalent, but we will not need this.

Proposition 3.1. *Let X and Y be two discrete random variables with finite sets of possible values, $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then, for every $\delta > 0$, there exists an $\varepsilon > 0$ such that if X and Y are ε-dependent, then $I(X;Y) \leq \delta$.*

Proof. We will give a direct proof of the statement. Assume X, Y are ε-dependent for some $\varepsilon > 0$. Let $p_{X,Y}$, p_X and p_Y denote the joint and marginal distributions of X and Y. Then, by the definition of ε-dependence, for $a \in \mathcal{X}$, $b \in \mathcal{Y}$ we have

$$p_{X,Y}(a,b) \leq p_X(a)p_Y(b) + \varepsilon \ ,$$
$$\frac{p_{X,Y}(a,b)}{p_X(a)p_Y(b)} \leq 1 + \frac{\varepsilon}{p_X(a)p_Y(b)} \ .$$

Using this and the fact that $\log_2(1+x) \leq \frac{x}{\ln 2}$ for all x, we can write

$$\begin{aligned}
I(X;Y) &= \sum_{a\in\mathcal{X}}\sum_{b\in\mathcal{Y}} p_{X,Y}(a,b)\log_2\left(\frac{p_{X,Y}(a,b)}{p_X(a)p_Y(b)}\right) \\
&\leq \sum_{a,b}(p_X(a)p_Y(b)+\varepsilon)\log_2\left(1+\frac{\varepsilon}{p_X(a)p_Y(b)}\right) \\
&\leq \sum_{a,b}(p_X(a)p_Y(b)+\varepsilon)\frac{1}{\ln 2}\frac{\varepsilon}{p_X(a)p_Y(b)} \\
&\leq \frac{1}{\ln 2}\text{card}(\mathcal{X})\text{card}(\mathcal{Y})\max_{a,b}\left\{\varepsilon+\frac{\varepsilon^2}{p_X(a)p_Y(b)}\right\} \\
&= \frac{1}{\ln 2}\text{card}(\mathcal{X})\text{card}(\mathcal{Y})\left(\varepsilon+\frac{\varepsilon^2}{\min_{a,b}\{p_X(a)p_Y(b)\}}\right)
\end{aligned}$$

Thus, $I(X;Y) \leq \delta(\varepsilon)$, where $\delta(\varepsilon) \coloneqq \frac{1}{\ln 2}\text{card}(\mathcal{X})\text{card}(\mathcal{Y})\left(\varepsilon + \frac{\varepsilon^2}{\min_{a,b}\{p_X(a)p_Y(b)\}}\right)$. Now, given $\delta > 0$, solve the quadratic equation $\delta(\varepsilon) = \delta$ for ε. Note that in a quadratic equation of the form $a\varepsilon^2 + b\varepsilon = \delta$ with $b, \delta > 0$, one of the solutions will always be positive. □

We are now able to state the generalization of Theorem 2.1.

As before, we let Γ be a countably infinite set.

Theorem 3.1. *Let $\underline{X} := (X_n)_{n\in\Gamma}$ be a family of random variables with values from a finite set M. Consider an event $E \subseteq M^\Gamma$, and suppose that for every finite set $J \subseteq \Gamma$ and every $\varepsilon > 0$, the event E has a (not necessarily injective) positional symmetry $\pi : \Gamma \to \Gamma$ such that*

(i) *X_J and $X_{\pi(J)}$ are ε-dependent,*

(ii) *The families $\underline{X}$ and $\underline{X}_\pi := (X_{\pi(n)})_{n\in\Gamma}$ have the same distribution.*

Then, $\mathbb{P}\left((X_n)_{n\in\Gamma} \in E\right) \in \{0,1\}$.

Before proving the theorem, we will state and prove two useful lemmas. We start with a generalization of O'Connell's Inequality to almost independent random variables.

Lemma 3.2 (O'Connell's Inequality for Almost Independent Random Variables)**.** *Let $\delta > 0$. Let $V, W_1, W_2, \ldots, W_n$ be discrete random variables and suppose that $I(W_i;(W_1, W_2, \ldots, W_{i-1})) < \delta$ for every index $i = 2, 3, \ldots, n$. Then*

$$H(V) > \sum_{i=1}^{n} I(V; W_i) - n\delta$$

Proof. Using the definition of mutual information on $I(W_i;(W_1, W_2, \ldots, W_{i-1})) < \delta$, we get that $H(W_i \mid (W_1, W_2, \ldots, W_{i-1})) > H(W_i) - \delta$ for every $i = 2, 3, \ldots, n$.

$$\begin{aligned}
H(V) &\geq H(V) - H(V \mid (W_1, W_2, \ldots, W_n)) \\
&= H(W_1, W_2, \ldots, W_n) - H(W_1, W_2, \ldots, W_n \mid V) \\
&= \sum_{i=1}^{n} H(W_i \mid W_1, W_2, \ldots, W_{i-1}) + \sum_{i=1}^{n} H(W_i \mid V, W_1, W_2, \ldots, W_{i-1}) \\
&> \sum_{i=1}^{n} [H(W_i) - \delta] - \sum_{i=1}^{n} H(W_i \mid V) \\
&= \sum_{i=1}^{n} [H(W_i) - H(W_i \mid V)] - n\delta \\
&= \sum_{i=1}^{n} I(V; W_i) - n\delta
\end{aligned}$$

□

Lemma 3.3. *In the setting of the theorem: Let $J \subseteq \Gamma$ be finite. For every $n \in \mathbb{N}$ and $\delta > 0$, there exist $J_1 = J, J_2, \ldots, J_n$ disjoint finite subsets of Γ such that,*

(i) *$I(X_{J_i}; (X_{J_1}, X_{J_2}, \ldots, X_{J_{i-1}})) < \delta$ for every $i = 1, 2, \ldots, n$*

(ii) *$I(\mathbb{1}_E(\underline{X}); X_{J_i}) = I(\mathbb{1}_E(\underline{X}); X_J)$ for every $i = 1, 2, \ldots, n$*

Proof. In light of Proposition 3.1, the hypothesis of the theorem can be restated as follows:

- For every finite $J \subseteq \Gamma$ and $\delta > 0$, the event E has an injective positional symmetry $\pi : \Gamma \to \Gamma$ such that

  (i') $I(X_J, X_{\pi(J)}) < \delta$,

  (ii') The families $\underline{X}$ and $\underline{X}_\pi$ have the same distribution.

Let $\delta > 0$ be arbitrary. We proceed by induction on n:

- Base step:

  For $n = 2$, we start with a injective positional symmetry π_1 corresponding to $W_1 := J_1$ and δ. Then, $J_2 := \pi_1(J_1)$ is such that $I(X_{J_1}; X_{J_2}) < \delta$ and $I(\mathbb{1}_E(\underline{X}); X_{J_1}) = I(\mathbb{1}_E(\underline{X}); X_{J_2})$ because, every $a \in \{0,1\}$ and $b \in M$,

$$\begin{aligned}\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_1} = b) &= \mathbb{P}(\mathbb{1}_E(\underline{X}_{\pi_1}) = a, X_{J_2} = b) && \text{(because } \underline{X}_{\pi_1} \sim \underline{X}\text{)} \\ &= \mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_2} = b) && \text{(because } \pi_1 \text{ is a symmetry of } E\text{)}\end{aligned}$$

- Inductive step:

  For $n > 1$, we assume the statement is true for $n-1$, and we have $J_1 = J, J_2, \dots, J_{n-1}$ with the desired properties. We choose a positional symmetry π_n corresponding to $W_{n-1} := \bigcup_{i=1}^{n-1} J_i$ and $\delta > 0$. Let $J_n := \pi_n(J_1)$.

  Then, $I(X_{J_n}; X_{W_{n-1}}) \leq I(X_{\pi_n(W_{n-1})}; X_{W_{n-1}}) < \delta$ by the data processing inequality and the choice of π_n. Furthermore, $I(\mathbb{1}_E(\underline{X}); X_{J_n}) = I(\mathbb{1}_E(\underline{X}); X_{J_1})$ because, for every $a \in \{0,1\}$ and $b \in M$,

$$\begin{aligned}\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_1} = b) &= \mathbb{P}(\mathbb{1}_E(\underline{X}_{\pi_n}) = a, X_{J_n} = b) && \text{(because } \underline{X}_{\pi_n} \sim \underline{X}\text{)} \\ &= \mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_n} = b) && \text{(because } \pi_n \text{ is a symmetry of } E\text{)}\end{aligned}$$

This proves the lemma. □

We are now able to prove the theorem.

Proof of Theorem 3.1. To prove Theorem 3.1, proceed similarly to that of Theorem 2.1.

Let $J \subseteq \Gamma$ be an arbitrary finite set. Given $n \in \mathbb{N}$ and $\delta > 0$, choose $J_1 = J, J_2, \dots, J_n$ as in Lemma 3.3. Then, using Lemma 3.2, we have

$$\begin{aligned}1 \geq H(\mathbb{1}_E(\underline{X})) &> \sum_{i=1}^{n} I(\mathbb{1}_E(\underline{X}); X_{J_i}) - n\delta \\ &= nI(\mathbb{1}_E(\underline{X}); X_J) - n\delta\end{aligned}$$

Therefore, $I(\mathbb{1}_E(\underline{X}); X_J) \leq \frac{1}{n} + \delta$. Letting first $n \to \infty$ and then $\delta \to 0$, it follows that $I(\mathbb{1}_E(\underline{X}); X_J) = 0$, that is, $\mathbb{1}_E(\underline{X})$ is independent of X_J. Since J is arbitrary, from Lemma 1.2 we find that $\mathbb{1}_E(\underline{X})$ is independent of the entire family $\underline{X}$. In particular, $\underline{X} \in E$ is independent of itself, which implies $\mathbb{P}(\underline{X} \in E) \in \{0,1\}$. □

As a corollary of Theorem 3.1, we obtain the following corollary for families of random variables with the Kolmogorov property. An event $V \subseteq M^\Gamma$ is called a *cylinder* event if there exists a finite set $A \subseteq \Gamma$ such that whether a sequence $(a_n)_{n\in\Gamma}$ belongs to V or not is determined by $(a_n)_{n\in A}$.

Definition 11 (Kolmogorov property)**.** A family $(X_n)_{n\in\Gamma}$ of random variables with values from a finite set M is said to have the *Kolmogorov property* (or has *short-range correlations*) if for every cylinder event $V \subseteq M^\Gamma$ and every $\varepsilon > 0$, there exists a finite set $B \subseteq \Gamma$ such that for every event $W \subseteq M^\Gamma$ determined by coordinates outside B, we have

$$\left|\mathbb{P}\big((X_n)_{n\in\Gamma} \in V \cap W\big) - \mathbb{P}\big((X_n)_{n\in\Gamma} \in V\big)\mathbb{P}\big((X_n)_{n\in\Gamma} \in W\big)\right| < \varepsilon\,.$$

(In other words, V and W are ε-dependent.)

The families with the Kolmogorov property are precisely those for which every tail event is deterministic [5, Prop. 7.9].

Here are some examples of families with the Kolmogorov property:

(a) Let $(X_n)_{n\in\mathbb{N}}$ be an irreducible and aperiodic Markov chain with a finite state space M. Then, $(X_n)_{n\in\mathbb{N}}$ has the Kolmogorov property [1].

(b) Every extremal Gibbs random field has the Kolmogorov property [5, Thm. 7.7].

Corollary 3.3.1. *Let $(X_n)_{n\in\Gamma}$ be a family of random variables with values from a finite set M that has the Kolmogorov property. Consider an event $E \subseteq M^\Gamma$, and suppose that for every finite set $J \subseteq \Gamma$, the event E has a positional symmetry $\pi : \Gamma \to \Gamma$ such that $\pi(J) \cap J = \varnothing$ and $(X_n)_{n\in\Gamma}$ and $(X_{\pi(n)})_{n\in\Gamma}$ have the same distribution. Then, $\mathbb{P}\left((X_n)_{n\in\Gamma} \in E\right) \in \{0, 1\}$.*

Proof. Choose B such that $J \subseteq B$ and ε so that any F outside B is ε-dependent on J. Now consider $\pi(B)$, which is disjoint of B. $\pi(J) \subseteq \pi(B)$, then it is outside B. Therefore, $X_{\pi(J)}$ and X_J are ε-dependent. Then by Theorem 3.1, the corollary is proved. □

3.2 Relaxing the Condition of Identical Distribution

In Theorem 3.1, the positional symmetries were required to preserve the distribution of the family $(X_n)_{n\in\Gamma}$ (Condition (ii)). In this section, we will relax this condition by allowing the distributions of $(X_n)_{n\in\Gamma}$ and $(X_{\pi(n)})_{n\in\Gamma}$ to be "close" rather than exactly the same.

We will define this notion within the theorem.

Theorem 3.4. *Let $\underline{X} := (X_n)_{n\in\Gamma}$ be a family of random variables with values from a finite set M. Consider an event $E \subseteq M^\Gamma$, and suppose that for every finite set $J \subseteq \Gamma$ and every $\varepsilon > 0$ and every $\delta > 0$, the event E has a positional symmetry $\pi : \Gamma \to \Gamma$ such that*

(i) *X_J and $X_{\pi(J)}$ are ε-dependent,*

(ii) *The families $\underline{X}$ and $\underline{X}_\pi = (X_{\pi(n)})_{n\in\Gamma}$ have almost the same distribution, i.e.*

$$\forall a, b, \left|\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_J = b) - \mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{\pi(J)} = b)\right| < \delta$$

Then, $\mathbb{P}\left((X_n)_{n\in\Gamma} \in E\right) \in \{0, 1\}$.

Similarly to the previous theorem, we will need Lemma 3.2 in addition to an extended version of Lemma 3.3, for which we need the following additional lemma.

Lemma 3.5. *Let X, Y, X', Y' be discrete random variables taking values from finite sets. For every $\kappa > 0$, there exists a $\delta > 0$ such that if for every a, b, we have $|\mathbb{P}(X = a, Y = b) - \mathbb{P}(X' = a, Y' = b)| < \delta$, then $|I(X, Y) - I(X', Y')| < \kappa$.*

Proof. Observe that the mutual information

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mathbb{P}(X = x, Y = y) \log \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)}$$

is a continuous function of the joint distribution $p_{X,Y}(x, y) := \mathbb{P}(X = x, Y = y)$ of X and Y because it is a composition of elementary continuous functions. □

We now prove a modified version of Lemma 3.3 with relaxed conditions.

Lemma 3.6. *In the setting of the theorem: Let $J \subseteq \Gamma$ be finite. For every $n \in \mathbb{N}$, $\varepsilon' > 0$, and $\kappa > 0$, there exist $J_1 = J, J_2, \ldots, J_n$ disjoint finite subsets of Γ such that,*

(i) *$I(X_{J_i}; (X_{J_1}, X_{J_2}, \ldots, X_{J_{i-1}})) < \varepsilon'$ for every $i = 1, 2, \ldots, n$*

(ii) *$|I(\mathbb{1}_E(\underline{X}); X_{J_i}) - I(\mathbb{1}_E(\underline{X}); X_J)| < \kappa$ for every $i = 1, 2, \ldots, n$*

Proof. In light of Theorem 3.4, the hypothesis of the theorem can be restated as follows:

- For every finite $J \subseteq \Gamma$, $\varepsilon' > 0$, and $\delta > 0$, the event E has an injective positional symmetry $\pi : \Gamma \to \Gamma$ such that

  (i') $I(X_J, X_{\pi(J)}) < \varepsilon'$,

  (ii') The families $(X_n)_{n\in\Gamma}$ and $(X_{\pi(n)})_{n\in\Gamma}$ have almost the same distribution, i.e.

$$\forall a, b, |\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_J = b) - \mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{\pi(J)} = b)| < \delta$$

Let $\varepsilon' > 0$ and $\kappa > 0$ be arbitrary and δ corresponding to κ from Lemma 3.5. We proceed by induction on n:

- Base step:

  For $n = 2$, we start with a positional symmetry π_1 corresponding to $W_1 := J_1$, ε', and δ. Then, $J_2 := \pi_1(J_1)$ is such that $I(X_{J_1}; X_{J_2}) < \varepsilon'$ and $|I(\mathbb{1}_E(\underline{X}); X_{J_1}) - I(\mathbb{1}_E(\underline{X}); X_{J_2})| < \kappa$ by Lemma 3.5 because, every $a \in \{0,1\}$ and $b \in M^J$,

$$|\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_1} = b) - \mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_2} = b)| \\ = |\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_1} = b) - \mathbb{P}(\mathbb{1}_E(\underline{X}_{\pi_1}) = a, X_{J_2} = b)| < \delta \,.$$

- Inductive step:

  For $n > 1$, we assume the statement is true for $n - 1$, and we have $J_1 = J, J_2, \ldots, J_{n-1}$ with the desired properties. We choose an injective positional symmetry π_n corresponding to $W_{n-1} := \bigcup_{i=1}^{n-1} J_i$, $\varepsilon' > 0$, and $\delta > 0$. Let $J_n := \pi_n(J_1)$.

  Then, $I(X_{J_n}; X_{W_{n-1}}) \leq I(X_{W_{n-1}}, X_{\pi_n(W_{n-1})}) < \varepsilon'$ by the data processing inequality and the choice of π_n. Furthermore, $|I(\mathbb{1}_E(\underline{X}); X_{J_n}) - I(\mathbb{1}_E(\underline{X}); X_{J_1})| < \kappa$ by Lemma 3.5 because, for every $a \in \{0,1\}$ and $b \in M$,

$$|\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_1} = b) - \mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_n} = b)| \\ = |\mathbb{P}(\mathbb{1}_E(\underline{X}) = a, X_{J_1} = b) - \mathbb{P}(\mathbb{1}_E(\underline{X}_{\pi_n}) = a, X_{J_n} = b)| < \delta \,.$$

This proves the lemma. □

We can finally prove the theorem with both conditions relaxed.

Proof of Theorem 3.4. Let $J \subseteq \Gamma$ be an arbitrary finite set. Given $n \in \mathbb{N}$, $\varepsilon' > 0$, and $\kappa > 0$, choose $J_1 = J, J_2, \ldots, J_n$ as in Lemma 3.6. Then, using Lemma 3.2, we have

$$1 \geq H(\mathbb{1}_E(\underline{X})) > \sum_{i=1}^{n} I(\mathbb{1}_E(\underline{X}); X_{J_i}) - n\varepsilon' \\ > nI(\mathbb{1}_E(\underline{X}); X_J) - n\varepsilon' - n\kappa$$

Therefore, $I(\mathbb{1}_E(\underline{X}); X_J) \leq \frac{1}{n} + \varepsilon' + \kappa$. Letting $n \to \infty$, $\varepsilon' \to 0$, and $\kappa \to 0$ (in that order), it follows that $I(\mathbb{1}_E(\underline{X}); X_J) = 0$, that is, $\mathbb{1}_E(\underline{X})$ is independent of X_J. Since J is arbitrary, from Lemma 1.2 we find that $\mathbb{1}_E(\underline{X})$ is independent of the entire family $\underline{X}$. In particular, $\underline{X} \in E$ is independent of itself, which implies $\mathbb{P}(\underline{X} \in E) \in \{0,1\}$. □

4 Approximate zero-one laws for finite collections? Two examples

In all previous sections, we studied infinite sequences of random variables. Our goal in this section is to look at what happens when we consider finite collections instead.

First, here are some useful inequalities which will be used in the proofs in this section.

Proposition 4.1 (Inequalities).

- *(Chebyshev's Inequality) Let X be a random variable with finite mean μ and finite variance σ^2. For every $k > 0$, we have*

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq 1/k^2$$

- *(Chernoff Bounds) Let $X_1, ..., X_n$ be i.i.d. Bernoulli random variables with parameter p, and let $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\overline{X}_n \geq p + \varepsilon\right) \leq 2^{-D(p+\varepsilon \| p)n} ,$$
$$\mathbb{P}\left(\overline{X}_n \leq p - \varepsilon\right) \leq 2^{-D(p-\varepsilon \| p)n} ,$$

where $D(q \parallel p) := q \log_2 \frac{q}{p} + (1-q)\log_2 \frac{1-q}{1-p}$ is the Kullback–Leibler divergence between Bernoulli distributions with parameters q and p.

- *(Pinsker's Inequality; Lemma 11.6.1 in [4])*

$$D(q \parallel p) \geq \frac{2(q-p)^2}{\ln 2}$$

We will look at two examples, in which we will consider an indicator function of an event regarding i.i.d. Bernoulli random variables, $g(X_1, X_2, ..., X_n)$, with the symmetry of exchangeability (which would make the event deterministic if there were infinitely many random variables), and study the probability distribution of that function.

We will notice in the first example that the function $\mathbb{P}(g(X_1, X_2, ..., X_n) = 1)$ behaves similarly to the below graph, suggesting that the symmetry may be the cause of the sharp transition from 0 to 1. However, we observe different results in the second example.

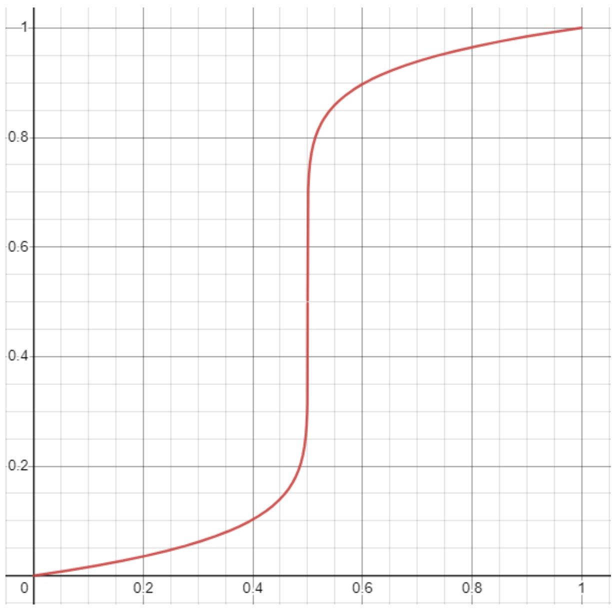

Figure 3: Sharp transition in the probability of $g(X_1, X_2, \ldots, X_n) = 1$.

4.1 Majority Element

The first example we will discuss is that of the majority element. A majority element of a sequence $a_1, a_2, ..., a_n$ is an element m such that it occurs at least $\frac{n}{2}$ times.

In our example, we consider a sequence of i.i.d. Bernoulli random variables with parameter p, $X_1, X_2, ..., X_n$. We define $g : \{0,1\}^n \to \{0,1\}$

$$g_n(a_1, a_2, \ldots, a_n) := \mathrm{maj}(a_1, a_2, \ldots, a_n) = \begin{cases} 1 & \text{if } \overline{a}_n \geq \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\overline{a}_n := (a_1 + a_2 + \cdots + a_n)/n$. Note that the event $g_n(X_1, \ldots, X_n) = 1$ is exchangeable, since permuting any number of elements will not affect the sum or average.

We describe the behavior of $\mathbb{P}_p[g_n(X_1, \ldots, X_n) = 1]$ in the following proposition.

Proposition 4.2. *In the above context,*

(i) *For every $\varepsilon > 0$, we have:*

- *If $p < \frac{1}{2} - \varepsilon$, then $\mathbb{P}_p[g_n(X_1, \ldots, X_n) = 1] < \delta$.*
- *If $p > \frac{1}{2} + \varepsilon$, then $\mathbb{P}_p[g_n(X_1, \ldots, X_n) = 1] > 1 - \delta$.*

where $\delta := \mathrm{e}^{-2\varepsilon^2 n}$.

(ii) *For $p = \frac{1}{2}$, we have $\frac{1}{2} - \delta' \leq \mathbb{P}_{1/2}[g_n(X_1, \ldots, X_n) = 1] \leq \frac{1}{2} + \delta'$, where $\delta' := \frac{e}{2\pi\sqrt{n}}$.*

This proposition gives us bounds on the distribution of the majority function g_n in terms of p, in particular, on $\varphi(p) := \mathbb{P}_p[g_n(X_1, \ldots, X_n) = 1]$ over the interval $[0,1]$. We point out that φ is a polynomial, hence it is continuous, and that it is monotone non-decreasing, having the shape of the graph in Figure 3. A simulation of the transition of φ over $[0,1]$ is shown in Figure 4. Part (i) implies that, given $\delta > 0$, the function φ transitions from being smaller than δ to being larger than $1 - \delta$ within a window of width at most $2\varepsilon = \sqrt{\frac{-2\ln\delta}{n}}$. Part (ii) says that $\varphi(\frac{1}{2})$ is approximately $\frac{1}{2}$ for large n.

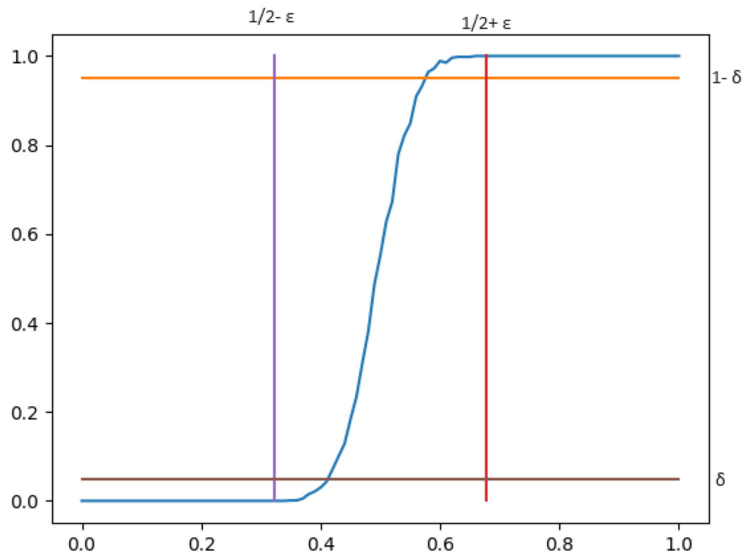

Figure 4: Simulation results of the transition in the probability of $g(X_1, X_2, \ldots, X_n) = 1$ (see Proposition 4.2) as a function of p (the probability to obtain 1 for an individual random variable) with $\varepsilon = 0.1766$ and $\delta = 0.05$ bounds. For a given p, in order to approximate the probability of $g(X_1, X_2, \ldots, X_n) = 1$ with $n = 100$, we have repeated the experiment 1000 times and computed the frequency of times in which $g(X_1, X_2, \ldots, X_n) = 1$. We calculated this frequency (probability) for 100 different values of p, starting with 0 and continuing with increments of 0.01. Clearly, it adheres to the model shown in Figure 3.

We now prove the proposition.

Proof of Proposition 4.2.

(i) Let $\varepsilon > 0$, and $p < \frac{1}{2} - \varepsilon$. Then

$$\begin{aligned}
\mathbb{P}_p[g_n(X_1, \ldots, X_n) = 1] &= \mathbb{P}_p\left[\overline{X}_n \geq \frac{1}{2}\right] \\
&= \mathbb{P}_p\left[\overline{X}_n \geq \left(\frac{1}{2} - \varepsilon\right) + \varepsilon\right] \\
&\leq \mathbb{P}_p[\overline{X}_n > p + \varepsilon] \\
&\leq 2^{-D(p+\varepsilon\|p)n} && \text{(by the Chernoff bounds)} \\
&\leq 2^{\frac{-2\varepsilon^2 n}{\ln 2}} && \text{(by Pinsker's inequality)} \\
&= \mathrm{e}^{-2\varepsilon^2 n} \\
&= \delta\,.
\end{aligned}$$

The proof of the other statement is almost identical.

(ii) Let $p = \frac{1}{2}$,

$$\mathbb{P}_{1/2}\left[\overline{X}_n < \frac{1}{2}\right] + \mathbb{P}_{1/2}\left[\overline{X}_n = \frac{1}{2}\right] + \mathbb{P}_{1/2}\left[\overline{X}_n > \frac{1}{2}\right] = 1$$

By symmetry, $\mathbb{P}_{1/2}\left[\overline{X}_n < \frac{1}{2}\right] = \mathbb{P}_{1/2}\left[\overline{X}_n > \frac{1}{2}\right]$. If n is odd, $\mathbb{P}_{1/2}\left[\overline{X}_n = \frac{1}{2}\right] = 0$, hence $\mathbb{P}_{1/2}\left[\overline{X}_n <\right.$

$\frac{1}{2}\Big] = \mathbb{P}_{1/2}\Big[\overline{X}_n > \frac{1}{2}\Big] = \frac{1}{2}$. If n is even, then we can write

$$\begin{aligned}\mathbb{P}_{1/2}\Big[\overline{X}_n > \frac{1}{2}\Big] &= \frac{1}{2} - \frac{1}{2}\mathbb{P}_{1/2}\Big[\overline{X}_n = \frac{1}{2}\Big] \\ &= \frac{1}{2} - \frac{1}{2}\binom{n}{\frac{n}{2}}\left(\frac{1}{2}\right)^n \\ &\geq \frac{1}{2} - \frac{1}{2}\cdot\frac{\mathrm{e}}{\pi}\cdot\frac{2^n}{\sqrt{n}}\left(\frac{1}{2}\right)^n \qquad \text{(using Stirling's approximation)} \\ &= \frac{1}{2} - \frac{\mathrm{e}}{2\pi\sqrt{n}}\end{aligned}$$

Hence, setting $\delta' := \frac{\mathrm{e}}{2\pi\sqrt{n}}$, we get what we need by symmetry. □.

4.2 Odd Number of 1's

In this example, we again consider a finite sequence of i.i.d. $X_1, X_2, ..., X_n$ of Bernoulli random variables with parameter p. We will look at the event that the sum of this sequence is odd.

For this, we define $g : \{0,1\}^n \to \{0,1\}$

$$g_n(a_1, a_2, \ldots, a_n) := a_1 + a_2 + ... + a_n \pmod 2$$

Note that the function takes value 1 when there is an odd number of 1's in the sequence, and value 0 when there is an even number of 1's in the sequence. Also, the event that we have an odd number of 1's is exchangeable because it only depends on the sum of the random variables.

Proposition 4.3. *In the above context,*

$$q_n = \mathbb{P}(g_n(X_1, X_2, ..., X_n) = 1) = \frac{1}{2} - \frac{1}{2}\cdot(1-2p)^n$$

Note that as n grows q_n goes to $\frac{1}{2}$ exponentially fast.

Proof. Using the law of total probability, we can write

$$\begin{aligned}q_{n+1} &= \mathbb{P}\big(g_{n+1}(X_1, X_2, ..., X_{n+1}) = 1\big) \\ &= \mathbb{P}(g_n(X_1, X_2, ..., X_n) = 1)\cdot\mathbb{P}(X_{n+1} = 0) + \mathbb{P}(g_n(X_1, X_2, ..., X_n) = 0)\cdot\mathbb{P}(X_{n+1} = 1) \\ &= q_n\cdot(1-p) + (1-q_n)\cdot p \\ &= q_n - q_n\cdot p + p - q_n\cdot p \\ q_{n+1} &= q_n\cdot(1-2p) + p\end{aligned}$$

Let $r_n = q_n - \frac{1}{2}$, then

$$r_{n+1} + \frac{1}{2} = \left(r_n + \frac{1}{2}\right)\cdot(1-2p) + p$$

or

$$r_{n+1} = r_n\cdot(1-2p)\ ,$$

which gives

$$r_n = r_1\cdot(1-2p)^{n-1}\ .$$

Rewriting this again in terms of q_n, and using the initial condition $q_1 = p$, we obtain

$$q_n = \frac{1}{2} - \frac{1}{2}\cdot(1-2p)^n\ .$$

as claimed. □

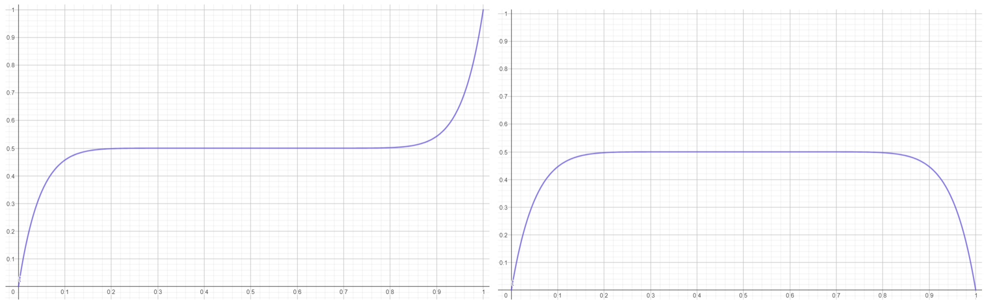

Figure 5: We observe a different graph for this example. The graph of $\varphi(p) \coloneqq \mathbb{P}_p\big(g_n(X_1, X_2, \ldots, X_n)\big)$ resembles the one to the left when n is odd, and the one to the right when n is even.

4.3 Concluding Comments

In an attempt to generalize the result of Section 4.1, we consider partitions of the interval $[0, 1]$ into disjoint sets A and B, and we define

$$g_n(a_1, a_2, \ldots, a_n) \coloneqq \begin{cases} 1 & \text{if } \overline{a}_n \in A, \\ 0 & \text{if } \overline{a}_n \in B. \end{cases}$$

If we attempt to graph $\mathbb{P}_p[g_n(X_1, X_2, \cdots, X_n) = 1]$ with respect to p, we should expect the graph to be pulled towards 1 in the interior of A and towards 0 in the interior of B. The larger n is, the clearer this behavior is. However, the behavior can be tricky to understand on the boundary of A and B, especially if these boundaries are infinite. We leave it as an open question to more carefully study the behavior in more general cases like infinite boundary or other families of examples.

References

[1] David Blackwell and David Freedman. The tail σ-field of a Markov chain and a theorem of Orey. *Annals of Mathematical Statistics*, 35(3), 1964. doi: 10.1214/aoms/1177703284.

[2] Peter J. Cameron. The random graph revisited. In C. Casacuberta, R. M. Miró-Roig, J. Verdera, and S. Xambó-Descamps, editors, *European Congress of Mathematics (Barcelona, July 10–14, 2000)*, volume 201 of *Progress in Mathematics*, pages 267–274. Birkhäuser, 2001. doi: 10.1007/978-3-0348-8268-2_15.

[3] Peter J. Cameron. The random graph. In R. Graham, J. Nešetřil, and S. Butler, editors, *The Mathematics of Paul Erdős II*. Springer, 2nd edition, 2013. doi: 10.1007/978-1-4614-7254-4_22.

[4] Thomas Cover and Joy Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 1991.

[5] Hans-Otto Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter, 2nd edition, 2011.

[6] C. Ward Henson. A family of countable homogeneous graphs. *Pacific Journal of Mathematics*, 38(1): 69–83, 1971. doi: 10.2140/pjm.1971.38.69.

[7] Neil O'Connell. Information-theoretic proof of the Hewitt-Savage zero-one law. Technical Report HPL-BRIMS-2000-18, Hewlett-Packard Laboratories, 2000.