# Random Sparse Bit Strings at the Threshold of Adjacency

(Extended Abstract appeared in STACS '98)

Joel H. Spencer[1] and Katherine St. John[2]

[1] Courant Institute, New York University, New York, NY 10012
[2] Department of Mathematics, Santa Clara University, Santa Clara, CA 95053-0290

## 1 Abstract

We give a complete characterization for the limit probabilities of first order sentences over sparse random bit strings at the threshold of adjacency. For strings of length $n$, we let the probability that a bit is "on" be $\frac{c}{\sqrt{n}}$, for a real positive number $c$. For every first order sentence $\phi$, we show that the limit probability function:

$$f_\phi(c) = \lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \text{ has the property } \phi]$$

(where $U_{n,\frac{c}{\sqrt{n}}}$ is the random bit string of length $n$) is infinitely differentiable. Our methodology for showing this is in itself interesting. We begin with finite models, go to the infinite (via the almost sure theories) and then characterize $f_\phi(c)$ as an infinite sum of products of polynomials and exponentials. We further show that if a sentence $\phi$ has limiting probability 1 for some $c$, then $\phi$ has limiting probability identically 1 for every $c$. This gives the surprising result that the almost sure theories are identical for every $c$.

## 2 Introduction

Expressibility is a central question in computer science. Over classes of ordered finite structures, membership in a complexity classes is often equivalent to the expressibility of the desired set in a given logic. For example, Immerman [6] showed that the expressibility in transitive closure logic is equivalent to NLOGSPACE, and Fagin [4] proved $\Sigma_1^1$ captures NPTIME. The characterizations of logics and the limit probabilities of their sentences over ordered structures could shed light on issues in complexity theory.

We focus on the class of ordered structures with a single unary predicate– that is, bit strings. Besides being a natural class to consider, logic, over bit strings, offers a useful tool to characterize the languages accepted by finite state automata. If we let the alphabet of our automata be $\{0, 1\}$, then the words in the language are bit strings. First order logic captures exactly the plus free regular languages, while monadic second order logic expresses the regular languages (see [7] and chapter 5 of [3]). We discuss the behavior of first order logic over random sparse bit strings and raise some interesting open problems about monadic second order logic in Section 5.

If we allow all bit strings to occur with equal probability, then for every first order sentence, $\phi$,

$$\lim_{n\to\infty} \frac{\#\text{ of models of size } n \text{ with property } \phi}{\text{total } \#\text{ of models of size } n}$$

converges. We can focus on bit strings where a small number of bits are on by allowing a bit to be "on" with probability $p(n)$ that depends on $n$, the length of the string. The random unary predicate $U_{n,p}$ is a probability space over predicates $U$ on $[n] = \{1, \ldots, n\}$ with the probabilities determined by $\Pr[U(x)] = p(n)$, for $1 \leq x \leq n$, and the events $U(x)$ are mutually independent over $1 \leq x \leq n$. $U_{n,p}$ is also called the *random bit string*. To see the correspondence, write each structure as a sequence of 0's and 1's, with the $i$th element in the sequence a 1 if and only if $U(i)$ holds in the structure. For example, if $n = 5$ and the unary predicate holds only on the least element, we write: [10000].

In [11], Shelah and Spencer showed that for every such sentence $\phi$ and for $p(n) \ll n^{-1}$ or $n^{-1/k} \ll p(n) \ll n^{-1/(k+1)}$, there is convergence for the limit probability. That is, there exists a constant $a_\phi$ such that

$$\lim_{n\to\infty} \Pr[U_{n,p} \models \phi] = a_\phi \tag{1}$$

Dolan [2] showed that for $p(n) \ll n^{-1}$ and $n^{-1} \ll p(n) \ll n^{-1/2}$ for every $\phi$, $a_\phi = 0$ or 1 in Equation 1. This stronger convergence is called a Zero-One Law for $U_{n,p}$. Dolan also showed that the Zero-One Law does not hold for $n^{-1/k} \ll p(n) \ll n^{-1/(k+1)}$, $k > 1$.

In [14], we examine the random sparse bit strings with probability $p(n) = c/n$ and give a finer analysis than convergence. For this choice of $p$, we have the limit probabilities of $\phi$ are either

$$\sum_{i=1}^{i=m} e^{-c}\frac{c^{t_i}}{t_i!} \qquad \text{or} \qquad 1 - \sum_{i=1}^{i=m} e^{-c}\frac{c^{t_i}}{t_i!}$$

for some (possibly empty) sequence of positive integers $t_1, \ldots, t_m$. We achieve a simpler characterization for $p = c/n$ due to the simpler underlying structures. Other interesting structures that have also been examined in this fashion are random graphs (without order) with edge probability $p(n) = c/n$ and $p(n) = \ln n/n + c/n$ (see the work of Lynch, Spencer, and Thoma: [9], [10], and [13]).

For each real constant $c$, let $S_c$ be the almost sure theory of the linear ordering with $p(n) = \frac{c}{\sqrt{n}}$. That is,

$$S_c = \{\phi \mid \lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \phi] = 1\}$$

Let $T_1$ be the almost sure theory for $n^{-1} \ll p(n) \ll n^{-1/2}$ and $T_2$ be the almost sure theory for $n^{-1/2} \ll p(n) \ll n^{-1/3}$. By [2], we have that $T_1$ is a complete theory. We characterize the theories at the threshold of adjacency, namely the $S_c$'s. These, in some sense, lie between $T_1$ and $T_2$. For each first order formula, $\phi$, we define the function:

$$f_\phi(c) = \lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \phi]$$

where $c$ ranges over the real, positive numbers. We show that $f_\phi(c)$ is infinitely differentiable. Moreover, we show:

**Theorem 1** *For every first-order sentence $\phi$, $f_\phi(c)$ is*

$$\sum_m e^{-tc^2}\frac{c^m}{m!}g(t,m)$$

*where for every fixed $t$, $g(t,m)$ is the sum of the product of polynomials and exponentials in $m$.*

In fact, the above theorem holds for any $t > 2^r$ where $r$ is the quantifier rank of the sentence $\phi$. The function $g(t,m)$ counts the number of equivalence classes of models (with respect to the maximum number, $m$, of pairs of 1's of distance at most $t$ from one another) that satisfy $\phi$.

We further show that if a sentence $\phi$ has limiting probability 1 for some $c$, then $\phi$ has limiting probability identically 1 for every $c$. This gives the surprising result that the almost sure theories are identical for every $c$.

**Theorem 2** *Let $S = \bigcap_c S_c$ be the intersection of all the almost sure theories. Then, for every real, positive $c$, $S_c = S$.*

To prove these theorems, we look first at the countable models of the almost sure theories (for more on this, see [12]). Let $\mathcal{U} \models S_c$ be such a model. Each of these models satisfy a set of basic axioms $\Delta$ (defined in Section 3). Let $\mathbf{a} = (a_0, \ldots, a_{m-1})$ be a finite sequence of non-negative integers, representing the distance pairs of 1's occur apart. We show, using Ehrenfeucht-Fraisse games, that for every first order sentence, $\phi$, $\Delta \cup \{\sigma_{\mathbf{a},t}\} \models \phi$ and $\Delta \cup \{\sigma_{\mathbf{a},t}\} \models \neg\phi$ where $t > 2^r$ for $r$ the quantifier rank of $\phi$, and $\sigma_{\mathbf{a},t}$ is the first-order sentence that states the 1's that occur within $t$ of one another are exactly those specified in the finite sequence $\mathbf{a}$ in exactly the order in $\mathbf{a}$. For example, if $\mathbf{a} = (2,0)$, then $\sigma_{\mathbf{a},3}$ would say that the first pair of 1's, that occur within three of one another, occur with exactly 2 0's in between, and the next (and only other) pair of 1's that occurs within three are adjacent. We show that

$$q_{\mathbf{a},t} = \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \sigma_{\mathbf{a},t}] = e^{-tc^2}\frac{c^{2m}}{m!}$$

To show our results, we "transfer" to another language. Let $\mathcal{L}$ be the language of random bit strings, and for each positive integer $t$, let $\mathcal{L}'_t$ be the language with equality, linear ordering, and a $t$-valued function, $d$. We will define a function $\mathcal{F}_t$ that takes the countable models of the almost sure theories, in the language $\mathcal{L}$, to structures in the language, $\mathcal{L}'_t$. Roughly, $\mathcal{F}_t$ takes each model $\mathcal{U}$ to an ordered set of positive integers that characterize the pairs of 1's that occur within $t$ of one another. If this set is finite, then it is the (unique) sequence of positive integers $\mathbf{a}$ such that $\mathcal{U} \models \sigma_{\mathbf{a},t}$. For example, if $\mathcal{U} \models \sigma_{(2,0),3}$, then $\mathcal{F}_t(\mathcal{U})$ is the two element structure $[0,1]$ such that $d(0) = 2$ and $d(1) = 0$. Just as we wrote the unary structures as bit strings, we can write the models of $\mathcal{L}'_t$ as ordered sequences of $\{0, \ldots, t-1\}$. So, for our example, we would write $[2,0]$. We show:

**Theorem 3** *Fix a real, positive constant, c. Let $\mathcal{U}_1$,$\mathcal{U}_2$ be models of an almost sure theory $S_c$. If $\mathcal{F}_t(\mathcal{U}_1)$ and $\mathcal{F}_t(\mathcal{U}_2)$ agree on $\mathcal{L}'_t$-sentences of quantifier rank at most t (that is, $\mathcal{F}_t(\mathcal{U}_1) \equiv_t \mathcal{F}_t(\mathcal{U}_2)$), then $\mathcal{U}_1$ and $\mathcal{U}_2$ agree on all $\mathcal{L}$-sentences of quantifer rank at most t (that is, $\mathcal{U}_1 \equiv_t \mathcal{U}_2$).*

These facts give the desired form for $f_\phi(c)$ in Theorem 1 and are used to show Theorem 2.

## 3 Examples

To give some intuition about what these models and theories look like, we begin with an informal discussion of the almost sure theories $T_1$ and $T_2$. For $T_1$, we have $\frac{1}{n} \ll p(n) \ll \frac{1}{\sqrt{n}}$, and almost surely isolated 1's occur. To see this, let $A_i$ be the event that $U(i)$ holds, $X_i$ be the random indicator variable, and $X = \sum_i X_i$, the total number of 1's that occur (i.e. the total number of elements for which the unary predicate holds). Then, $E(X_i) = p(n)$, and by linearity of expectation,

$$E(X) = \sum_i E(X_i) = np(n).$$

We have $E(X) \to \infty$ as $n \to \infty$. Since all the events are independent, $\mathrm{Var}[X] \leq E[X]$. By the Second Moment Method (see [1], chapter 4 for details),

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{E[X]^2} \leq \frac{E[X]}{E[X]^2} = \frac{1}{E[X]} \to 0$$

Thus, $\Pr[X > 0] \to 1$. So, almost surely, arbitrarily many 1's occur. We can write this in first order logic as a schema of sentences $\alpha_r$, each of which states "there is at least $r$ 1's":

$$\alpha_r : (\exists x_1 \dots x_r)(x_1 < x_2 < \cdots < x_r \wedge U(x_1) \wedge \cdots \wedge U(x_r))$$

Each of these 1's occurs arbitrarily far apart. To see this, let $B_i$ be the event that $i$ and $i+1$ are 1's, let $Y_i$ be its random indicator variable, and $Y = \sum_i Y_i$. Then, $E(Y_i) = \Pr[B_i] = p^2$ and $E(Y) \sim np^2 \to 0$. So, almost surely, 1's occur, but no 1's occur adjacent in the order. If, for each $r > 0$, we let $C_{i,r}$ be the event that $i$ and $i+r$ are 1's and $C_r = \sum_i C_{i,r}$, we can show, by similar argument, that $C_r \to 0$. This works for any fixed $r$, so, the 1's that do occur are isolated from one another by arbitrarily many 0's. So, almost surely, the schema of sentences $\beta_r$ that state that "between every pair of 1's there is $r$ 0's" hold:

$$\begin{aligned}\beta_r : (\forall x_1, x_2)[(&U(x_1) \wedge U(x_2) \wedge x_1 < x_2) \to \\ &(\exists y_1, \dots, y_r)(\neg U(y_1) \wedge \ldots \wedge \neg U(y_r) \wedge x_1 < y_1 < \cdots < y_r < x_2)]\end{aligned}$$

Thus, for every $r$, $\alpha_r, \beta_r \in T_1$, the almost sure theory.

The almost sure theory also contains sentences about the ordering. Since every $U_{n,p}$ is linearly ordered with a minimal and maximal element, the first-order sentences that state these properties are in $T_1$, $T_2$, and each $S_c$. Let $\Gamma_l$ be the order axioms for the linear theory, that is, the sentences:

$$
\begin{array}{l}
(\forall xyz)[(x \leq y \wedge y \leq z) \rightarrow x \leq z] \\
(\forall xy)[(x \leq y \wedge y \leq x) \rightarrow x = y] \\
(\forall x)(x \leq x) \\
(\forall xy)(x \leq y \vee y \leq x)
\end{array}
$$

The following sentences guarantee that there is a minimal element and a maximal element:

$$
\begin{array}{l}
\mu_1 : (\exists x \forall y)(x \leq y) \\
\mu_2 : (\exists x \forall y)(x \geq y)
\end{array}
$$

There is also a minimal and maximal 1, which can be stated as:

$$
\begin{array}{l}
\mu_1' : (\exists x)(\forall y)[(U(x) \wedge U(y)) \rightarrow (x \leq y)] \\
\mu_2' : (\exists x)(\forall y)[(U(x) \wedge U(y)) \rightarrow (x \geq y)]
\end{array}
$$

Further, every element, except the maximal element, has a unique successor under the ordering, and every element, except the minimal element, has a unique predecessor. This can be expressed in the first-order language as:

$$
\begin{array}{l}
\eta_1 : (\forall x)[(\forall y)(x \geq y) \vee (\exists y \forall z)((x \leq y \wedge \neg(x = z)) \rightarrow y \leq z) \\
\eta_2 : (\forall x)[(\forall y)(x \leq y) \vee (\exists y \forall z)((x \geq y \wedge \neg(x = z)) \rightarrow y \geq z)
\end{array}
$$

As $n \rightarrow \infty$, the number of elements also goes to infinity. To capture this, we add for each positive $r$ the axiom:

$$\delta_r : (\exists x_1 \ldots x_r)(x_1 < x_2 < \cdots < x_r)$$

For $n \geq r$, $U_{n,p} \models \delta_r$. Thus, for every $U_{n,p}$,

$$U_{n,p} \models \Gamma_l \wedge \mu_1 \wedge \mu_2 \wedge \eta_1 \wedge \eta_2 \wedge \delta_1 \wedge \delta_2 \wedge \ldots \wedge \delta_n$$

We also have that for every $j$, there exists an $n$ such that

$$U_{n,p} \models \alpha_l \wedge \ldots \wedge \alpha_j \wedge \beta_1 \wedge \ldots \wedge \beta_j$$

Let $\Delta = \{\Gamma_l, \mu_1, \mu_2, \eta_1, \eta_2, \bigwedge_r \alpha_r, \bigwedge_r \delta_r\}$. Then $\Delta \subset T_1, T_2$ and for each $c > 0$, $\Delta \subset S_c$. The set of sentences, $\Sigma \cup \{\bigwedge_r \beta_r\}$, axiomatizes $T_1$. This follows from an Ehrenfeucht-Fraisse game argument (see [12] for more details).

For countable models of $T_1$, we cannot have a single infinite chain, since all the 1's must be isolated. So, we must have infinitely many chains, ordered like the integers (called **Z**-chains) that contain a single 1 with an infinite increasing chain of 0's at the beginning and an infinite decreasing chain of 0's at the end. Between these can be any number of **Z**-chains that contain no 1's. Call any **Z**-chain that contains a 1 **distinguished**. For any distinguished **Z**-chain, except the maximal distinguished chain, almost surely, there's a least distinguished **Z**-chains above it (this follows from the discreteness of the finite models). In other words, every distinguished **Z**-chain, except the maximal 1, has a distinguished successor **Z**-chain. This rules out a "dense" ordering of the distinguished **Z**-chains and leads to a "discreteness" of 1's, similar to the discreteness of elements we encountered above. It says nothing about **Z**-chains without 1's– those could have any countable order type they desire. So, the simplest model is pictured in Figure 1.

$$[00\cdots)(\cdots 00100\cdots)\underbrace{(\cdots 00100\cdots)}_{\text{"a } \mathbf{Z}\text{-chain"}}\cdots \quad \cdots(\cdots 00100\cdots)(\cdots 00100\cdots)(\cdots 00]$$

**Fig. 1.** A model of $T_1$

When $\frac{1}{\sqrt{n}} \ll p(n) \ll \frac{1}{\sqrt[3]{n}}$, almost surely isolated 1's occur, as well as more complicated occurrences of 1's. The more complicated occurrences, which we will refer to as level 2 occurrences, are $11, 101, 1001, \ldots, 10^r1\ldots$, where "$10^r1$" is an interval $[i, i+r+1]$ with $U(i)$, $U(i+r+1)$, and for each $1 \le j \le i+r$, $\neg U(i+j)$. Using the notation from above, note $E(Y) = np^2 \to \infty$ and $\Pr[Y > 0] \to 1$. By similar argument, we can also show that three 1's cannot occur "close" together. Again, the distinguished **Z**-chains (i.e. those that contain at least 1 1) in a model of $T_2$ cannot be dense. The argument above can be extended to give that for every $r, s > 0$, almost surely for any occurrence of $10^r1$, except the maximal 1, there exists a least occurrence of $10^s1$ above it. In between any two level 2 occurrences, we have arbitrarily many isolated 1's. These cannot be densely ordered since almost surely every 1 has a successor. So, these sequences cannot be densely ordered either. That leaves only the **Z**-chains without 1's. Since we have no way to say things about them, they can have any countable order type they wish. Further, every finite sequence of level 2 occurrences must occur. Since $T_2 \models \Delta$, any model of $T_2$ begins with an ascending chain of 0's. In fact, each model will begin with a model of $T_1$, followed by a level 2 occurrence. Which level 2 occurrence occurs first is not fixed. [12] gives more details about the countable models of the almost sure theories $T_1$ and $T_2$.

When $p = \frac{c}{\sqrt{n}}$ and using the notation from above, the expected number of 1's in $U_{n,\frac{c}{\sqrt{n}}}$ is $E(X) = n \cdot \frac{c}{\sqrt{n}} = c\sqrt{n} \to \infty$. The expected number of pairs of 1's in $U_{n,\frac{c}{\sqrt{n}}}$ is $E(Y) = n \cdot p^2 = n \cdot \frac{c^2}{n} = c^2$. In any countable model of the almost sure theory, $S_c$, we have infinitely many isolated 1's, and we also have a non-zero probability of pairs of 1's occurring close together. Figure 2 shows a possible model of $S_c$. Note that the model in Figure 2 has two level 2 occurrences, namely, of length 2 and of length 0.

Let $\mathbf{a} = (a_0, \ldots, a_{m-1})$ be a finite sequence of non-negative integers. Let $m = |\mathbf{a}|$ be the length of the sequence and $M = \max\{a_0, \ldots, a_{m-1}\}$ be the maximum value of the sequence $\mathbf{a}$. Then, for each $t > M$, we can define a first order sentence $\sigma_{\mathbf{a},t}$ that says the only pairs of 1's that occur within $t$ of one another are exactly the distance prescribed by $\mathbf{a}$ and in that order.

$$
\begin{array}{ccccc}
 & \vdots & \vdots & & \\
 & (\cdots 00100 \cdots) & (\cdots 00100 \cdots) & & \\
 & (\cdots 00100 \cdots) & (\cdots 00100 \cdots) & & \\
[00\cdots)(\cdots 010 \cdots)\cdots & (\cdots 00100100 \cdots) & (\cdots 001100 \cdots) & \cdots(\cdots 010 \cdots)(\cdots 00] \\
 & (\cdots 00100 \cdots) & (\cdots 00100 \cdots) & & \\
 & (\cdots 00100 \cdots) & (\cdots 00100 \cdots) & & \\
 & \vdots & \vdots & &
\end{array}
$$

**Fig. 2.** A model of $S_c$

For example, if $\mathbf{a} = (2,0)$, then $\mathbf{a}$ represents the level 2 occurrences "1001" and "11", occurring in that order. $\sigma_{\mathbf{a},3}$ is the sentence:

$$
\begin{array}{l}
(\exists x_1 x_2 x_3 x_4)[U(x_1) \wedge U(x_2) \wedge U(x_3) \wedge U(x_4) \wedge x_1 < x_2 < x_3 < x_4 \\
\wedge(\exists y_1 y_2)(x_1 < y_1 < y_2 < x_2 \wedge (\forall z)(x_1 < z < x_2 \rightarrow (y_1 = z \vee y_2 = z))) \\
\wedge(\forall z)(x_3 < z \rightarrow (x_4 = z \vee x_4 < z)) \\
\wedge(\forall w_1 w_2)(U(w_1) \wedge U(w_2) \wedge w_1 < w_2 \\
\wedge(\neg(w_1 = x_1) \wedge \neg(w_1 = x_3)) \rightarrow (\exists y_1 y_2 y_3 y_4)(x_1 < y_1 < y_2 < y_3 < y_4 < x_2))]
\end{array}
$$

which states that the only 1's occuring within three of one another are "1001" and "11," in that order.

**Definition 1** *Fix $t$ and $c$ and let $\mathcal{U}$ be a model of $S_c$ (that is, $\mathcal{U} \models S_c$. Let*

$$D = \{(x_i, a_i) \mid \mathcal{U} \models U(x_i) \wedge U(x_i + a_i + 1) \wedge \neg U(x_i + 1) \wedge \ldots \neg U(x_i + a_i) \text{ for } a_i < t\ \}$$

*and let $A = \{x \mid$ There exists $a,(x,a) \in D\}$. We define the function $\mathcal{F}_t$ from models of $S_c$ to models of $\mathcal{L}'_t$ as follows: Define $\mathcal{F}_t(\mathcal{U})$ as the structure $< A, \leq', d >$ where $\leq'$ is the order induced from $\mathcal{U}$ (that is, $x \leq' y$ in $\mathcal{F}_t(\mathcal{U})$ iff $x \leq y$ in $\mathcal{U}$) and $d$ is the $t$-valued function with the set $D$ as its graph (that is, $d(x) = a$ iff $(d,a) \in D$).*

For example, if $\mathcal{U} \models \sigma_{(2,0),3}$, then $\mathcal{F}_3(\mathcal{U})$ can be written as $[2,0]$.

## 4 The Results

We use several facts and theorems from [12]. First, we can view $\mathcal{U}$ as a sequence of models of $T_1$, separated by pairs of 1's occurring within $t$ of one another. That is:

**Theorem 4** *([12]) Fix a positive integer $t$ and a real positive constant $c$. Let $\mathcal{U} \models S_c$. Let $x_i$ be the element at which the pair of length $a_i < t$ begins in $\mathcal{U}$, $x_m$ be element at which the last pair of length $< t$ occurs, and* max *be the maximal element in $\mathcal{U}$. Then,*

$$< [0, x_0 - 1], \leq, U >, < [x_0 + a_0, x_1 - 1], \leq, U >, \ldots, < [x_m + a_m, \max], \leq, U >$$

*are models of the almost sure theory $T_1$.*

We also have that the almost sure theory $T_1$ is complete, which gives:

**Theorem 5** *([12]) Fix a positive integer $t$. Let $\mathcal{U}_1$ and $\mathcal{U}_2$ be models of $T_1$. Then, $\mathcal{U}_1 \equiv_t \mathcal{U}_2$.*

We can now prove the transfer theorem, Theorem 3. Recall Theorem 3 states that if $\mathcal{U}_1, \mathcal{U}_2 \models S_c$ and $\mathcal{F}_t(\mathcal{U}_1) \equiv_t \mathcal{F}_t(\mathcal{U}_2)$, then $\mathcal{U}_1 \equiv_t \mathcal{U}_2$. That is, we can "transfer" the winning strategy from the finite structures of $\mathcal{L}'_t$ to the corresponding infinite models of $\mathcal{L}$.

*Proof of Theorem 3:* Assume $\mathcal{U}_1, \mathcal{U}_2 \models S_c$. By assumption, $\mathcal{F}_t(\mathcal{U}_1) \equiv_t \mathcal{F}_t(\mathcal{U}_2)$. Thus, we have a winning strategy for the $t$-move EF game played on $\mathcal{F}_t(\mathcal{U}_1)$ and $\mathcal{F}_t(\mathcal{U}_2)$.

We need to show that this winning strategy can be used to give Duplicator a winning strategy for the $t$-move game on $\mathcal{U}_1$ and $\mathcal{U}_2$. We show this by induction on $q$, the number of moves remaining in the EF game on $\mathcal{U}_1$ and $\mathcal{U}_2$. We will play with two sets of pebbles, one for the actual game on $\mathcal{U}_1$ and $\mathcal{U}_2$, and a "shadow" set for the game on $\mathcal{F}_t(\mathcal{U}_1)$ and $\mathcal{F}_t(\mathcal{U}_2)$.

Without loss of generality, assume Spoiler plays on the element $x$ in $\mathcal{U}_1$. Let $i$ be the index of the closest pair of 1's within distance t of one another to $x$ (if $x$ is infinitely far from the pair above and below it, let $i$ be the index of the one below it). Place the Spoiler's shadow pebble on $i$ in $\mathcal{F}_t(\mathcal{U}_1)$. By hypothesis, we have a winning strategy for Duplicator for any game on $\mathcal{F}_t(\mathcal{U}_1)$ and $\mathcal{F}_t(\mathcal{U}_2)$. Let $j$ be the move corresponding to $i$ and place Duplicator's shadow pebble on $j$. Returning to the actual game, roughly Duplicator plays on the same relative distance from the pair $b_j$ as Spoiler did from $a_i$.

In more detail, we need to take in account distance (up to $2^q$, where $q$ is the number of moves remaining) from the endpoints, the other placed pebbles, and the pairs of 1's of distance less than $t$. Here, we appeal to Theorem 4. From it, we have that every interval $\mathcal{M}_1 =< [x_i + a_i, x_{i+1} - 1], \leq, U_1 >$ and $\mathcal{M}_2 =< [y_j + b_j, y_{j+1} - 1], \leq, U_2 >$ are models of $T_1$, where $x_i$ is the element that begins the pair $a_i$, $y_j$ is the element that begins the pair $b_j$, $U_1$ is the restriction of the unary predicate of $\mathcal{U}_1$, and $U_2$ is the restriction of the unary predicate of $\mathcal{U}_2$. By Theorem 5, Duplicator has a winning strategy for any $q$-move EF game played on $\mathcal{M}_1$ and $\mathcal{M}_2$. So, for any move of Spoiler in such a interval, Duplicator has a winning move. This gives $\mathcal{U}_1 \equiv_t \mathcal{U}_2$. $\dashv$

### Using the Janson Inequalities

The Janson Inequalities (see [1], chapter 8 for more details) says that events that are "mostly" independent sometimes have probability "nearly equal" to the truly independent case. We will use these inequalities to give the limiting probability that a sequence of pairs $\mathbf{a}$ are the only occuring of length up to $t$.

**Lemma 1** *Let $c$ be a real positive constant, $\mathbf{a} = (a_0, \ldots, a_{m-1})$, a finite sequence of positive integers and $t$ a positive integer such that $t > 2^{\max\{a_1,\ldots,a_m,m\}}$. Then,*

$$q_{\mathbf{a},t} = \lim_{n\to\infty} Pr[U_{n,\frac{c}{\sqrt{n}}} \models \sigma_{\mathbf{a},t}] = e^{-tc^2} \frac{c^{2|\mathbf{a}|}}{|\mathbf{a}|!}$$

The above lemma gives that the likelihood of a sequence occurring depends solely on its length and becomes less likely as the sequence length increases. As a corollary, we have:

**Lemma 2** *Let $A$ be an ordered, countable set of positive integers from $\{0,\ldots,t-1\}$. Then the probability, as $n \to \infty$, that $\mathcal{U}_{n,c/\sqrt{n}}$ contains the pairs listed in $A$ is zero.*

With this in mind, we can focus on the limit probabilities of finite sequences, since those are the only sequences that make positive contributions to the limit probability.

**Definition 2** *For each first-order sentence $\phi$ and $t > qr(\phi)$, let*

$$M(\phi,t) = \{\mathbf{a} \mid S \cup \{\sigma_{\mathbf{a},t}\} \models \phi\} \text{ and } f_{\phi,t}(c) = \sum_{\mathbf{a}\in M(\phi,t)} q_{\mathbf{a},t}.$$

Recall that $f_\phi(c) = \lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \phi]$. We show that for every $s,t > qr(\phi)$, $f_{\phi,t} = f_{\phi,s}$. That is, the value of $f_{\phi,t}$ is fixed for sufficiently large $t$.

**Lemma 3** *For a finite sequence $\mathbf{a} = (a_0,\ldots,a_{m-1})$ with maximum value $M$ and for every $s > t > M$, let $\theta = \sigma_{\mathbf{a},M}$, then $f_{\theta,s}(c) = q_{\mathbf{a},t}$.*

Thus, excluding pairs up to $M$, the limit probability of $\mathbf{a}$ has the same limit probability for any $t > M$. It follows immediately:

**Corollary 1** *For every first order sentence $\phi$, and for every $s > t > qr(\phi)$, we have: $f_{\phi,s}(c) = f_{\phi,t}(c)$.*

### Proofs of the Theorems

*Proof of Theorem 1 :*

Let $\phi$ be a first-order sentence. By Lemma 1, for each $\mathbf{a}$, $t$:

$$q_{\mathbf{a},t} = \lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \sigma_{\mathbf{a},t}] = e^{-tc^2}\frac{c^{2m}}{m!}$$

where $m = |\mathbf{a}|$. Then, for fixed $t$,

$$\begin{aligned}\lim_{m_0\to\infty} \textstyle\sum_{m=0}^{m_0} \sum_{|\mathbf{a}|=m, a_i<t} q_{\mathbf{a},t} &= \textstyle\sum_{m=0}^{\infty} \sum_{|\mathbf{a}|=m, a_i<t} q_{\mathbf{a},t} \\ &= e^{-tc^2} \textstyle\sum_{m=0}^{\infty} \frac{(tc^2)^m}{m!} = 1\end{aligned}$$

So, for any positive $\epsilon > 0$, there exists $m_0 > t$ such that

$$\left| 1 - \sum_{m=0}^{m_0} \sum_{|\mathbf{a}|=m, a_i<t} q_{\mathbf{a},t} \right| < \epsilon$$

Let $\beta_{s_0} = \neg \bigvee_{m \leq m_0} \bigvee_{\sigma_{\mathbf{a},t} || \mathbf{a}|=m, a_i<t}$. We have

$$\lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \beta_{s_0}] = \left| 1 - \sum_{m=0}^{m_0} \sum_{|\mathbf{a}|=m, a_i<t} q_{\mathbf{a},t} \right| < \epsilon$$

**Claim 1** *$\phi$ has limiting probability* $\sum_{\mathbf{a}\in M(\phi,t)} q_{\mathbf{a},t}$ *where $t > qr(\phi)$.*

Using the claim, $f_\phi(c) = \lim_{n\to\infty} \Pr[U_{n,\frac{c}{\sqrt{n}}} \models \phi] = \sum_{\mathbf{a}\in M(\phi)} e^{-c} \frac{c^{2m}}{m!}$ where $m = |\mathbf{a}|$. ⊣

*Proof of Theorem 2:* By definition, $S \subseteq \bigcap_c S_c$. To show $\bigcap_c S_c \subseteq S$, assume $\phi \in S_c$ for some $c$. Then, $f_\phi(c) = 1$. So, for $t > 2^{qr(\phi)}$:

$$1 = f_\phi(c) = = e^{-tc^2} \sum_{m=0}^{\infty} (\sum_{|\mathbf{a}|=m, \mathbf{a}\in M(\phi,t)} \tfrac{c^{2m}}{m!})$$

This happens if and only if

$$\sum_{m=0}^{\infty} \left( \sum_{|\mathbf{a}|=m, \mathbf{a}\in M(\phi,t)} \frac{c^{2m}}{m!} \right) = e^{tc^2}$$

This occurs, if and only if, for each $m$, the number of $\mathbf{a} \in M(\phi)$ of length $m$ is $t^m$. But, for each $m$, this is the total number possible in $M(\phi, t)$, so, we must have that every $\mathbf{a} \in M(\phi)$. Thus, for every possible finite sequence $\mathbf{a}$, $S \cup \sigma_{\mathbf{a},t} \models \phi$. So, $f_\phi(c)$ is constantly 1, and $\phi \in S_c$ for every $c$. Therefore, $S = \bigcap_c S_c$. ⊣

## 5 Future Work

The work of [11] and [12] characterize the almost sure theories and their countable models for $p(n) \ll n^{-1}$ and $n^{-1/k} \ll p(n) \ll n^{-1/(k+1)}$ for $k \geq 1$. In [14] and this paper, we fill the "gaps" between these theories by characterizing the almost sure theories of $U_{n,\frac{c}{n}}$ and $U_{n,\frac{c}{\sqrt{n}}}$ and giving the form of the function $f_\phi(c)$ for each first order sentence $\phi$. Monadic second order logic is more expressive than first order logic over bit strings. For example, "evenness" can be expressed in monadic second order logic but not in first order logic. Is there a characterization for the limit probabilities of monadic second order logic over random sparse bit strings with $p = c/n$?

Let $\mathcal{L}_P$ be the language with the basic operations of addition, ordering, and the unary predicate. What happens to the limit probabilities of sentences over this extended language? Lynch [8] gave sufficient conditions on the unary predicates to be indistinguishable under sentences of quantifier rank less than $k$, for a fixed $k$ over the natural numbers. Grädel [5] linked subclasses of Presburger arithmetic (the first order theory of the model $< \mathbf{N}, +, \leq >$) to the polynomial time hierarchy, and related the truth value of sentences of quantifier rank $k$ to the truth on an initial segment of $\mathbf{N}$ whose length is dependent on $k$. These techniques might be useful on this question and related ones in random sequences.

## References

1. Noga Alon, Joel H. Spencer, and Paul Erdős. *The Probabilistic Method*. John Wiley and Sons, Inc, New York, 1992.
2. Peter Dolan. A zero-one law for a random subset. *Random Structures and Algorithms*, 2:317–326, 1991.
3. Heinz-Dieter Ebbinghaus and Jorg Flum. *Finite Model Theory*. Springer, Berlin, 1995.
4. Ronald Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In *Complexity of Computation, SIAM-AMS Proceedings*, pages 43–73, 1974.
5. Erich Graedel. Subclasses of Presburger arithmetic and the polynomial-time hierarchy. *Theoretical Computer Science*, 56:289–301, 1988.
6. Neil Immerman. Expressibility as a complexity measure: results and directions. In *Second Structure in Complexity Conference*, pages 194–202. Springer, 1987.
7. Harry R. Lewis and Christos H. Papadimitriou. *Elements of the theory of computation*. Prentice-Hall, Englewood Cliffs, NJ, 1981.
8. James F. Lynch. On sets of relations definable by addition. *Journal of Symbolic Logic*, 47:659–668, 1982.
9. James F. Lynch. Probabilities of sentences about very sparse random graphs. *Random Structures and Algorithms*, 3:33–54, 1992.
10. Saharon Shelah and Joel H. Spencer. Can you feel the double jump. *Random Structures and Algorithms*, 5(1):191–204, 1994.
11. Saharon Shelah and Joel H. Spencer. Random sparse unary predicates. *Random Structures and Algorithms*, 5(3):375–394, 1994.
12. Joel H. Spencer and Katherine St. John. Random unary predicates: Almost sure theories and countable models. Submitted for publication, 1997.
13. Joel H. Spencer and Lubos Thoma. On the limit values of probabilities for the first order properties of graphs. Technical Report 97-35, DIMACS, 1997.
14. Katherine St. John. Limit probabilities for random sparse bit strings. *Electronic Journal of Combinatorics*, 4(1):R23, 1997.

This article was processed using the LaTeX macro package with LLNCS style