Limit Probabilities for Random Sparse Bit Strings

Katherine St. John *
Department of Mathematics
University of Pennsylvania
Philadelphia, Pennsylvania 19104
stjohn@math.upenn.edu

Submitted: May 12, 1997; Accepted: October 14, 1997

Abstract

Let n be a positive integer, c a real positive constant, and $p(n) = c/n$. Let $U_{n,p}$ be the random unary predicate under the linear order, and S_c the almost sure theory of $U_{n,\frac{c}{n}}$. We show that for every first-order sentence ϕ:

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

is an infinitely differentiable function. Further, let $S = \bigcap_c S_c$ be the set of all sentences that are true in every almost sure theory. Then, for every $c > 0$, $S_c = S$.

(Mathematical Reviews Subject Classification: 03C13, 60F20, 68Q05)

1 Introduction

Let n be a positive integer, and $0 \le p(n) \le 1$. 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 $1 \le x \le n$, and the events $U(x)$ are mutually independent over $1 \le x \le n$. $U_{n,p}$ is also called the random bit string.

Let ϕ be a first-order sentence in the language with linear order and the unary predicate. In [7], Shelah and Spencer showed that for every such sentence ϕ and for $p(n) \ll n^{-1}$ or $n^{-1/k} \ll p(n) \ll n^{-1/(k+1)}$, there exists a constant a_ϕ such that

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

*Current Address: Department of Mathematics, Santa Clara University, Santa Clara, CA 95053-0290, kstjohn@scu.edu.

(Note that "$U_{n,p} \models \phi$" means that $U_{n,p}$ has property ϕ. See Section 2 for this and other definitions.)

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

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

Let T_0 be the almost sure theory for $p(n) \ll n^{-1}$, and T_1 be the almost sure theory for $n^{-1} \ll p(n) \ll n^{-1/2}$. By the work of Dolan [2], $U_{n,p}$ satisfies the 0-1 law for $p(n) \ll n^{-1}$ and $n^{-1} \ll p(n) \ll n^{-1/2}$ (that is, for every ϕ, $a_\phi = 0$ or 1 in Equation 1). This gives that T_0 and T_1 are complete theories. Dolan also showed that the 0-1 Law does not hold for $n^{-1/k} \ll p(n) \ll n^{-1/(k+1)}$, $k > 1$.

In this paper, we will characterize the theories between T_0 and T_1, namely the S_c's. For each first-order formula, ϕ, define the function:

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

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 ϕ, $f_\phi(c)$ is either*

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

for some finite (possibly empty) sequence of positive integers $t_1, \ldots, t_m$.

Let $S = \bigcap_c S_c$ be the set of all sentences that are true in every almost sure theory. We show:

Theorem 2 *For every real, positive c, $S_c = S$.*

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: [5], [6], and [9]). We achieve a simpler characterization of the limit probabilities than those for random graphs due to our underlying models.

To prove these theorems, we look first at the countable models of the almost sure theories (for more on this, see [8]). Let $\mathcal{M} \models S_c$ be such a model. Each of these models satisfy a set of basic axioms Δ (defined in Section 3). Further, we show, using Ehrenfeucht-Fraisse games, that $\Delta \cup \{\sigma_i\}$ is complete, where σ_i is the first-order sentence that states "there are exactly i elements for which the unary predicate holds." For every $\mathcal{M}$, there is an i such that $\mathcal{M} \models \Delta \cup \{\sigma_i\}$. For each first-order sentence ϕ, either ϕ follows from only a finite number of complete extensions, or the negation of ϕ, $\neg\phi$, follows from only a finite number of complete extensions. Let X be the number of elements for which the unary predicate holds in $U_{n,\frac{c}{n}}$. The $\Pr[X = i]$ has binomial distribution. These two facts give the desired form for $f_\phi(c)$ in Theorem 1 and are used to show Theorem 2.

In an effort to keep the paper self-contained and accessible, we have included many definitions and concepts that the expert in the field might wish to skip. Section 2 of this paper includes definitions from logic and finite model theory. To illustrate the definitions, we have included a section of Examples (Section 3). Section 4 includes the proofs of the results.

A note on notation: we will use lower case Greek letters for first-order sentences $(\phi, \psi, \ldots)$, upper case Greek letters for sets of sentences $(\Gamma, \Delta, \ldots)$, and lower case Roman letters to refer to elements in the universe $(i, j, \ldots)$.

2 Definitions

This section contains the definitions we need from first-order logic and finite model theory. A more thorough treatment of first-order logic can be found in Enderton [4], of finite model theory in Ebbinhaus and Flum [3], and of the probabilistic method in Alon, Spencer, and Erdős [1].

We concentrate on first-order logic over the basic operations $\{\leq, U, =\}$. That is, we are interested in sentences made up of $=$ (equality), $\leq$ (linear order), U (an unary predicate), the binary connectives $\vee$ (disjunction) and $\wedge$ (conjunction), $\neg$ (negation), and the first-order quantifiers $\exists$ (existential quantification) and $\forall$ (universal quantification). "First-order" refers to the range of the quantifiers– we only allow quantification over variables, not sets of variables. For example,

$$(\exists x)(\forall y)(x \leq y)$$

is a first-order sentence that expresses the property that there is a least element. The x and y are assumed to range over elements of the universe. A set of consistent sentences is often called a **theory**.

Our structures have an underlying set $[n] = \{1, \ldots, n\}$ with the basic operations: $=, \leq$ and U. Without loss of generality, we will interpret the ordering $\leq$ as the natural ordering on $[n]$. There are many choices for interpreting the unary predicate U over $[n]$ (2^n to be precise). We can view the structures as ordered sequences of 0's and 1's, where the ith element in the sequence is 1 if and only if $U(i)$. For example, if $n = 5$ and the unary predicate holds on the least element, then structure can be represented as [10000]. Let $\mathcal{M} =< [m], \leq, U >$, $\mathcal{M}_1 =< [m_1], \leq, U_1 >$, and $\mathcal{M}_2 =< [m_2], \leq, U_2 >$

be models where $\leq$ is a linear order on the universes (the underlying sets) of the structure, and U, U_1, and U_2 are unary predicates on the universes of their respective structure. We will say $\mathcal{M}$ **models** ψ (written: $\mathcal{M} \models \psi$) just in case the property ψ holds of $\mathcal{M}$. If for every model $\mathcal{M}$ we have $\mathcal{M} \models \Gamma$ implies $\mathcal{M} \models \psi$, where Γ is a (possibly empty) set of sentences, then we write $\Gamma \models \psi$ (pronounced "Γ models ψ"). For the particular ψ above, $\mathcal{M} \models \psi$ only if there is some element in $[m]$ which is less than or equal to every other element in $[m]$. Every $[m]$ has a least element (namely 1), so, $\mathcal{M} \models \psi$, and further, $\models \psi$.

While many things can be expressed using first-order sentences, many cannot. For example, there is no first-order sentence that captures the property that a structure has an universe with an even number of elements (see [3], p. 21). That is, there is no first-order sentence ϕ such that for every model $\mathcal{M} =< [m], \leq, U >$,

$$\mathcal{M} \models \phi \iff m \text{ is even}$$

One measure of the complexity of first-order sentences is the nesting of quantifiers. If a formula ϕ has no quantifiers, we say it has **quantifier rank** 0, and write $qr(\phi) = 0$. For all formulas, we define **quantifier rank** by induction:

- If $\phi = \phi_1 \vee \phi_2$ or $\phi = \phi_1 \wedge \phi_2$, then $qr(\phi) = \max(qr(\phi_1), qr(\phi_2))$.
- If $\phi = \neg\phi_1$, then $qr(\phi) = qr(\phi_1)$.
- If $\phi = \exists x \phi_1$ or $\phi = \forall x \phi_1$, then $qr(\phi) = qr(\phi_1) + 1$.

Definition 1 *For each t, two models $\mathcal{M}_1$ and $\mathcal{M}_2$ are* **equivalent** *(with respect to t), $\mathcal{M}_1 \equiv_t \mathcal{M}_2$ if they have the same truth value on all first-order sentences of quantifier rank at most t.*

The equivalence of structures under all first-order sentences of quantifier rank less than or equal to t is connected to the t-pebble games of Ehrenfeucht and Fraisse, described in [3]. Given two structures $\mathcal{M}_1$ and $\mathcal{M}_2$, $\mathcal{M}_1 \equiv_t \mathcal{M}_2$ if and only if the second player has a winning strategy for every t-pebble Ehrenfeucht-Fraisse game played on $\mathcal{M}_1$ and $\mathcal{M}_2$. We define the game below:

Definition 2 *The* **t-pebble Ehrenfeucht-Fraisse game (EF game)** *on $\mathcal{M}_1$ and $\mathcal{M}_2$ is a two-person game of perfect information. For the game, we have:*

- **Players:** *There are two players:*
  - *Player I, often called* Spoiler*, who tries to ruin any correspondence between the structures.*

  – *Player II, often called* Duplicator*, who tries to duplicate Spoiler's last move.*

- **Equipment:** *We have t pairs of pebbles and the two structures $\mathcal{M}_1$ and $\mathcal{M}_2$ as game boards.*
- **Moves:** *The players take turns moving. At the ith move, the Spoiler chooses a structure and places his ith pebble on an element in that structure. Duplicator then places her ith pebble on an element in the other structure.*
- **Winning:** *If after any of Duplicator's moves, the substructures induced by the pebbles are not isomorphic, then Spoiler wins. After both players have played t moves, if Spoiler has not won, then Duplicator wins.*

We say a player has a **winning strategy** for the t-pebble game on $\mathcal{M}_1$ and $\mathcal{M}_2$, if no matter how the opponent plays, the player can always win. These games form a powerful tool for showing theories are complete. A theory T is **complete** if for every sentence ϕ, either $T \models \phi$ or $T \models \neg\phi$. Equivalently, a theory T is complete if any two models of T satisfy the same first order sentences. For each q, we can show this for sentences of quantifier rank q by proving that Duplicator has a winning strategy for the q-move EF game on any two models of T. We will use this reduction to games to show the completeness of our theories S_c (see Lemma 1).

3 Examples

To give some intuition about what these models and theories look like, we begin with an informal discussion of T_0 and T_1. When $p(n) \ll n^{-1}$, almost surely the unary predicate never holds (i.e. no 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,

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

As n gets large, $E(X) \rightarrow 0$. Since $\Pr[X > 0] \leq E(X)$, almost surely, no 1's occur. This gives

$$\lim_{n\rightarrow\infty} \Pr[U_{n,p} \models (\exists x)U(x)] = 0.$$

The negation of this statement, $(\forall x)\neg U(x)$, almost surely is true. So, $(\forall x)\neg U(x)$ is in the almost sure theory T_0.

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_0, T_1, and each S_c. Let Γ_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:

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

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:

$$\begin{array}{ll}\eta_1 : & (\forall x)[(\forall y)(x \geq y) \vee (\exists y \forall z)((x \leq y \wedge x \neq z) \rightarrow y \leq z)] \\ \eta_2 : & (\forall x)[(\forall y)(x \leq y) \vee (\exists y \forall z)((x \geq y \wedge x \neq z) \rightarrow y \geq z)]\end{array}$$

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 $U_{n,p}$,

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

and $\Sigma = \{\Gamma_l, \mu_1, \mu_2, \eta_1, \eta_2, \bigwedge_r \delta_r\} \subset T_k$, for $k = 1, 2$, and for each c, $\Sigma \subset S_c$.

For T_0, the only additional axiom we need is $(\forall x)(\neg U(x))$. The simplest model of T_0 is an infinite decreasing chain of 0's followed by an infinite increasing chain of 0's (see Figure 1). More complicated models also satisfy $\Sigma \cup (\forall x)(\neg U(x))$, namely those with arbitrarily many copies of of chains of 0's, ordered like the integers (called "**Z**-chains"), with an infinite decreasing chain of 0's at the beginning and an infinite increasing chain of 0's at the end. The ordering of the **Z**-chains is not determined. It could be finite, infinite with discrete points, or it could be "dense." By the latter, we mean that between any 2 **Z**-chains, there's another.

When $n^{-1} \ll p(n) \ll n^{-1/2}$, almost surely isolated 1's occur. Using the notation above, we have $E(X) \rightarrow \infty$ as $n \rightarrow \infty$. Since all the events are independent, $\mathrm{Var}[X] \leq E[X]$. By the Second Moment Method (see [1], chapter 4 for details),

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

$$[000\cdots)\ (\cdots 000]$$

Figure 1: A model of T_0

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

Figure 2: A model of T_1

Thus, $\Pr[X > 0] \to 1$. 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.

For 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 **Z**-chains that contain a single 1. 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 2.

By the earlier discussion, we know that the basic axioms $\Sigma \subset T_1$. The only further axioms needed are those that guarantee arbitrarily many 1's occurring far apart and the "discreteness" of 1's. These axioms echo the basic axioms listed before, for each

$$[00\cdots)(\cdots 00100\cdots)(\cdots 00100\cdots)(\cdots 00]$$

Figure 3: A model of S_c

r we have:

$$\begin{array}{rl}
\mu_1' : & (\exists x)(\forall y)[(U(x) \wedge U(y)) \rightarrow (x \leq y)] \\
\mu_2' : & (\exists x)(\forall y)[(U(x) \wedge U(y)) \rightarrow (x \geq y)] \\
\delta_r' : & (\exists x_1 \ldots x_r)(x_1 < x_2 < \cdots < x_r \wedge U(x_1) \wedge \cdots \wedge U(x_r)) \\
\epsilon_r : & (\forall x_1, x_2)[(U(x_1) \wedge U(x_2) \wedge x_1 < x_2) \rightarrow \\
 & (\exists y_1, \ldots, y_r)(\neg U(y_1) \wedge \ldots \wedge \neg U(y_r) \wedge x_1 < y_1 < \cdots < y_r < x_2)]
\end{array}$$

These axioms, along with Σ, axiomatize T_1, which follows from an EF game argument.

When $p = c/n$, the expected number of 1's in $U_{n,\frac{c}{n}}$ is (using the notation from above):

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p(n) = n \cdot \frac{c}{n} = c.$$

In fact, $\Pr[X = t]$ has binomial distribution and the limiting probability is Poisson (see Lemma 3). So, with probability $\Pr[X = t] \rightarrow e^{-c}\frac{c^t}{t!}$, $U_{n,\frac{c}{n}}$ has exactly t 1's. In any countable model of the almost sure theory, S_c, the 1's occur arbitrarily far apart, as in models of T_1. A simple model of S_c, which occurs with probability $e^{-c}\frac{c^2}{2!}$ is in Figure 3. Let Δ be the axioms Σ along with the axiom schema ϵ_r for every r that guarantees the 1's are isolated (that is, $\Delta = \Sigma \cup \bigcup_r \{\epsilon_r\}$). For each S_c, $\Delta \subset S_c$.

4 The Results

For each natural number i, let σ_i be the first-order sentence that says "there exist exactly i 1's." So,

$$\begin{array}{rcl}
\sigma_1 & := & (\exists x)[U(x) \wedge (\forall y)(U(y) \rightarrow y = x)] \\
\sigma_{i+1} & := & (\exists x_1 x_2 \ldots x_{i+1})[U(x_1) \wedge U(x_2) \wedge \ldots \wedge U(x_{i+1}) \\
 & & \wedge\ x_1 < x_2 \wedge x_2 < x_3 \wedge \cdots x_i < x_{i+1} \\
 & & \wedge\ (\forall y)(U(y) \rightarrow (y = x_1 \vee y = x_2 \vee \cdots \vee y = x_{i+1}))]
\end{array}$$

Recall that we defined the set of sentences S to be all sentences that hold in every almost sure theory (that is, $S = \bigcap_c S_c$). Then:

Lemma 1 *For each i, $S \cup \{\sigma_i\}$ is complete.*

Proof: Note that the basic axioms Δ (defined in Section 3) are contained in S. So, it suffices to show that $\Delta \cup \{\sigma_i\}$ is complete (that is for every sentence ϕ, either $\Delta \cup \{\sigma_i\} \models \phi$ or $\Delta \cup \{\sigma_i\} \models \neg\phi$). We will show this is a complete theory by giving, for every t, a winning strategy for Duplicator for the t-move game on two models, $\mathcal{M}_1$ and $\mathcal{M}_2$ of $\Delta \cup \{\sigma_i\}$. The essence of the proof is that in our theories the 1's are spaced arbitrarily far apart. So, what matters in pebble placement is the distance to the closest 1. Since t pebbles can only tell distances of length $\leq 2^t$, we define the t-type of an interval to keep track of small distances from 1's.

Let the **t-type** of an interval $[a,b]$ to be (L,R,O,Z), with O the number of 1's in the interval; Z the number of 0's in the interval; L the minimal nonnegative number with $a+L$ a 1; R the minimal nonnegative number with $b-R$ a 1 – but if any of these numbers are not in the set $\{0,1,\ldots,2^t\}$ call them by a special symbol MANY_t. (That is, if the first 2^t+1 symbols of the interval are 0's then $L = \mathrm{MANY}_t$).

The strategy for Duplicator with t moves remaining and $x_1 < \ldots < x_s$ the moves already made on model $\mathcal{M}_1$; $x'_1 < \ldots < x'_s$ the moves already made on model $\mathcal{M}_2$ is to ensure that for all i intervals $[x_i, x_{i+1}]$, $[x'_i, x'_{i+1}]$ have the same t-type. To include the end intervals, assume Spoiler starts by playing the minimal and maximal elements of M to which Duplicator of course follows on M'. Since both $\mathcal{M}_1$ and $\mathcal{M}_2$ model σ_i, each has the same number of 1's occurring, the t-type of the initial moves are the same, namely $(\mathrm{MANY}_t, \mathrm{MANY}_t, \mathrm{MANY}_t, \mathrm{MANY}_t)$ for i sufficiently larger than t ($i > 3^t$ suffices).

We show that if $[a,b], [a',b']$ have the same t-type then for all $x \in [a,b]$ (Spoiler move) there exists $x' \in [a',b']$ (Duplicator move) with $[a,x], [a',x']$ having the same $(t-1)$-type and $[x,b], [x',b']$ also having the same $(t-1)$-type (similarly for every $x' \in [a',b']$). We proceed by induction on t, the number of moves remaining. For $t=1$, if $U(x)$, then we must have $O > 0$. So, there must be a $x' \in [a',b']$ such that $U(x')$. If $\neg U(x)$, then $Z > 0$ and there must be a $x' \in [a',b']$ such that $\neg U(x')$. Thus, Duplicator has a winning strategy for the game on intervals with the same 1-type and with 1 move remaining.

For $t > 1$, assume that $[a,b]$ and $[a',b']$ have the same t-type: (L,R,O,Z). Let (L_l, R_l, O_l, Z_l) be the $(t-1)$-type of $[a,x]$ and (L_r, R_r, O_r, Z_r) be the $(t-1)$-type of $[x,b]$. If $Z \neq \mathrm{MANY}_t$, then the lengths of the intervals $[a,b]$ and $[a',b']$ are equal and $\leq 2^t$. In this case, the t-type fully determines the occurrence and placement of any 1 in the interval (if one occurs). Let $x' = a' + x - a$. If $O = 0$, then $L = R = 0$, and both intervals are all 0's. If $O = 1$ (the only other possibility since $Z \neq \mathrm{MANY}_t$ and the 1's occur arbitrarily far apart), then the 1 occurs the exact same distance from x and x'. So, the resulting intervals $[a,x]$ and $[a',x']$, and $[x,b]$ and $[x',b']$ have the same t-types, and thus, $(t-1)$-types.

So, assume $Z = \mathrm{MANY}_t$, that is, the lengths of the intervals $[a,b]$ and $[a',b']$ are at least 2^t but may not be equal. If $x - a < 2^{t-1}$, let $x' = a' + x - a$. By construction,

$[a,x]$ and $[a',x']$ have the same length and the same number of 1's. If the number of 1's is zero, then $L_l = R_l = O_l = 0$ and $Z_l = x - a$ for both intervals. If the number of 1's is one (the only other possibility), then $L_l = L$, $R_l = \mathrm{MANY}_{t-1}$, $O_l = 1$, $Z_l = x - a - 1$ for both $[a,x]$ and $[a',x']$. Since $x - a < 2^{t-1}$, we have $b - x > 2^{t-1}$ and $Z_r = \mathrm{MANY}_{t-1}$. If $O_l = 0$, then the number of 1's in $[x,b]$ and $[x',b']$ is the same as the number of 1's in the original intervals (i.e. O). If the number of 1's is greater than 2^{t-1}, then $O_r = \mathrm{MANY}_{t-1}$. Otherwise, $O_r = O$. If $O_l = 1$, then the number of 1's in $[x,b]$ and $[x',b']$ is one less than that in the original intervals. So, $O_r = \mathrm{MANY}_{t-1}$ if $O - 1 > 2^{t-1}$, otherwise, $O_r = O - 1$. Thus, $[a,x]$ and $[a',x']$, and $[x,b]$ and $[x',b']$ have the same t-types, and thus, $(t-1)$-types. If $b - x < 2^{t-1}$ follows by a similar argument.

So, assume $a + 2^{t-1} \leq x \leq b + 2^{t-1}$. This gives that the length of both the leftside and rightside intervals is at least 2^{t-1}. Let $-2^{t-1} < y < 2^{t-1}$ be such that $U(x+y)$ if such a y exists. Let x' be such that $U(x'+y)$ and if $x+y$ is the ith 1 counting from the left for $i \leq 2^{t-1}$, then $x'+y$ is also the ith 1 counting from the left (such exists since both $[a,x]$ and $[a',x']$ have the same value for O). Similarly, if $x+y$ is the ith 1 counting from the right for $i \leq 2^{t-1}$, then $x'+y$ is also the ith 1 counting from the right. If neither of these hold, choose x' such that $x'+y$ is the ith 1 for $i > 2^t$. By construction, the resulting intervals will have the same values for $L_l, L_r, R_l, R_r, O_l, O_r$. The values for $Z_l = Z_r = \mathrm{MANY}_t$. So, $[a,x]$ and $[a',x']$, and $[x,b]$ and $[x',b']$ have the same t-types, and thus, $(t-1)$-types.

Lastly, assume $a + 2^{t-1} \leq x \leq b + 2^{t-1}$ but no y such that $U(x+y)$ and $-2^{t-1} < y < 2^{t-1}$ exists. Then x is at least 2^{t-1} from a, b, and every i such that $U(i)$. This gives $L_r = R_l = Z_l = Z_r = \mathrm{MANY}_{t-1}$. As before, the values of L_l and R_r depend on L and R (since they count the distance from endpoints that did not move). So, we only need for our choice of x' that it is at least 2^{t-1} from any occurrence of 1 and has the same value for O_l and O_r that $[a,x]$ and $[x,b]$ does. If $O_l < 2^{t-1}$, choose x' so that it occurs at least 2^{t-1} above the O_lth 1. If $O_r < 2^{t-1}$, then $[x',b]$ also has the same number of 1's since $O = O_l + O_r$ is the value for both $[a,b]$ and $[a',b']$. If $O_r = \mathrm{MANY}_{t-1}$, then, again $[x',b]$ also has the value $O_r = \mathrm{MANY}_{t-1}$. A similar argument works for $O_r < 2^{t-1}$. If both $O_l = O_r = \mathrm{MANY}_{t-1}$, then choose x' so that it occurs at least 2^{t-1} above the 2^{t-1}th 1 (such an x' exists, since $O = \mathrm{MANY}_t$). So, $[a,x]$ and $[a',x']$, and $[x,b]$ and $[x',b']$ have the same t-types, and thus, $(t-1)$-types.

Thus, $[a,x]$ and $[a',x']$ have the same $(t-1)$-types, as well as $[x,b]$ and $[x',b']$. By inductive hypothesis, Duplicator can win the $(t-1)$-move game played on $[a,x]$ and $[a',x']$, and on $[x,b]$ and $[x',b']$. Duplicator can win the t-move game on $[a,b]$ and $[a',b']$ by placing x' (x) according to the above strategy, and then following the strategy given by inductive hypothesis for the remaining $t-1$ moves. □

Definition 3 *For each first-order sentence ϕ, let $M(\phi) = \{i \mid S \cup \{\sigma_i\} \models \phi\}$.*

Lemma 2 *For each first-order ϕ, $M(\phi)$ is finite or co-finite.*

Proof: Let ϕ be a first-order sentence. Let $t = qr(\phi)$, the quantifier rank of ϕ. We claim that either $\max\{i \in M(\phi)\} < 3^t$ or $\max\{i \in M(\neg\phi)\} < 3^t$.

Assume not. Then, there exists $i, j > 3^t$ such that $i \in M(\phi)$ and $j \in M(\neg\phi)$. Let $\mathcal{M}_1 \models \Delta \cup \{\sigma_i\}$ and $\mathcal{M}_2 \models \Delta \cup \{\sigma_j\}$. $i \in M(\phi)$ also implies $\mathcal{M}_1 \models \phi$, and $j \in M(\neg\phi)$ implies $\mathcal{M}_2 \models \neg\phi$. Since i and j are sufficiently larger than t, the strategy in Lemma 1, gives that $\mathcal{M}_1$ and $\mathcal{M}_2$ agree on all first-order sentences of quantifier rank $\leq t$. This gives a contradiction since $\mathcal{M}_1$ and $\mathcal{M}_2$ disagree on ϕ. $\square$

Lemma 3 *Let c be a real positive constant and i a positive integer. Let X be the number of 1's occurring in $U_{n,c/n}$. Then $Pr[X = i]$ has a binomial distribution and $\lim_{n\to\infty} Pr[X = i]$ is Poisson.*

Moreover,

$$\lim_{n\to\infty} Pr[X = i] = \lim_{n\to\infty} Pr[U_{n,\frac{c}{n}} \models \sigma_i] = e^{-c}\frac{c^i}{i!}$$

Proof: For each i and c, the probability that a set of size i is all 1's is $(\frac{c}{n})^i$. The probability that the remaining $n - i$ vertices are all 0's is $(1 - \frac{c}{n})^{n-i}$. Since there are $\binom{n}{i}$ ways to choose such an i element set,

$$\Pr[X = i] \;=\; \binom{n}{i} (\tfrac{c}{n})^i (1 - \tfrac{c}{n})^{n-i}$$

As we let n gets large:

$$\begin{aligned} \lim_{n\to\infty} \Pr[X = i] &= \lim_{n\to\infty} \binom{n}{i} (\tfrac{c^i}{n^i})(1 - \tfrac{c}{n})^n (1 - \tfrac{c}{n})^{-i} \\ &= \lim_{n\to\infty} (\tfrac{n^i}{i!})(\tfrac{c^i}{n^i}) e^{-c} \cdot 1 \\ &= e^{-c}\tfrac{c^i}{i!} \end{aligned}$$

$\square$

Proofs of the Theorems

Proof of Theorem 1 : Let ϕ be a first-order sentence. Then, by Lemma 2, $M(\phi)$ is finite or co-finite. Assume first that $M(\phi)$ is finite (the co-finite case follows by similar argument on $M(\neg\phi)$), and there exists positive integers $t_1 < t_2 < \ldots < t_m$ such that $M(\phi) = \{t_1, \ldots, t_m\}$. By Lemma 3, for each positive integer, i,

$$q_i = \lim_{n\to\infty} \Pr[U_{n,\frac{c}{n}} \models \sigma_i] = e^{-c}\frac{c^i}{i!}$$

Then,

$$\lim_{s_0\to\infty}\sum_{s=0}^{s_0} q_s = \sum_{s=0}^{\infty} q_s = \sum_{s=0}^{\infty} e^{-c}\frac{c^s}{s!} = e^{-c}\sum_{s=0}^{\infty}\frac{c^s}{s!} = e^{-c}e^{c} = 1$$

So, for any positive $\epsilon > 0$, there exists $s_0 > t_m$ such that

$$\left|1 - \sum_{s=1}^{s_0} q_s\right| < \epsilon$$

Let $\beta_{s_0} = \neg\bigvee_{s\le s_0}\sigma_s$. We have

$$\lim_{n\to\infty}\Pr[U_{n,c/n}\models\beta_{s_0}] = 1 - \sum_{s=1}^{s_0} q_s < \epsilon \tag{2}$$

Claim 1 *ϕ has limiting probability $\sum_{s\in M(\phi)} q_s$.*

Proof of Claim: If suffices to show that $\phi\leftrightarrow\neg\bigvee_{t\in M(\phi)}\sigma_t$ has limiting probability 0. This breaks into the cases of $\phi\wedge\neg\bigvee_{t\in M(\phi)}\sigma_t$ and $\neg\phi\wedge\bigvee_{t\in M(\phi)}\sigma_t$. For the first, fix $\epsilon > 0$ and choose s_0 such that Equation 2 holds. Then $(\phi\wedge\neg\bigvee_{t\in M(\phi)}\sigma_t)\leftrightarrow\beta_{s_0}\vee\bigvee_{s\le s_0, s\notin M(\phi)}(\phi\wedge\sigma_s)$ is in the almost sure theory. $\bigvee_{s\le s_0, s\notin M(\phi)}(\phi\wedge\sigma_s)$ is a finite disjunction of sentences with limiting probability 0. So,

$$\begin{aligned}
\lim_{n\to\infty}\ & \Pr[U_{n,c/n}\models(\phi\wedge\neg\textstyle\bigvee_{t\in M(\phi)}\sigma_t)]\\
=\ & \lim_{n\to\infty}\Pr[(U_{n,c/n}\models\beta_{s_0}\vee\textstyle\bigvee_{s\le s_0, s\notin M(\phi)}(\phi\wedge\sigma_s))]\\
\le\ & \max\{\lim_{n\to\infty}\Pr[U_{n,c/n}\models\beta_{s_0}],\\
& \lim_{n\to\infty}\Pr[U_{n,c/n}\models\textstyle\bigvee_{s\le s_0, s\notin M(\phi)}(\phi\wedge\sigma_s)]\}\\
\le\ & \max\{\epsilon, 0\}\\
=\ & \epsilon
\end{aligned}$$

Since ϵ is arbitrary, $\lim_{n\to\infty}\Pr[U_{n,c/n}\models(\phi\wedge\neg\bigvee_{t\in M(\phi)}\sigma_t)] = 0$.

For the second case of $\neg\phi\wedge\bigvee_{t\in M(\phi)}\sigma_t$, we have the finite disjunction of $(\neg\phi\wedge\sigma_t)$, for $t\in M(\phi)$, each of which has limit probability 0. So,

$$\lim_{n\to\infty}\Pr[U_{n,c/n}\models\neg\phi\wedge\bigvee_{t\in M(\phi)}\sigma_t] = 0.$$

□

Using the claim,

$$\begin{aligned}
f_\phi(c) &= \lim_{n\to\infty}\Pr[U_{n,c/n}\models\phi]\\
&= \lim_{n\to\infty}\Pr[U_{n,c/n}\models(\sigma_{t_1}\vee\ldots\vee\sigma_{t_m})]\\
&= \textstyle\sum_{t\in M(\phi)}\lim_{n\to\infty}\Pr[U_{n,c/n}\models\sigma_t]\\
&= \textstyle\sum_{t\in M(\phi)} e^{-c}\frac{c^t}{t!}
\end{aligned}$$

□

Proof of Theorem 2: By definition, S is contained in S_c (since S is defined to be the intersection of the almost sure theories, $\bigcap_c S_c$). So, we need to show, that for every c, $S_c \subseteq S$. Fix a real positive constant d, and let ϕ be a first order sentence contained in S_d. Then, by definition, $f_\phi(d) = 1$. So, $M(\phi) \neq \emptyset$. Assume $M(\phi)$ is finite and $M(\phi) = \{t_1, \ldots, t_m\}$. By Theorem 1, for every positive c,

$$\begin{aligned} f_\phi(c) &= e^{-c} \textstyle\sum_{i=1}^{i=m} \frac{c^{t_i}}{t_i!} < e^{-c} \sum_{i=1}^{i=\infty} \frac{c^i}{i!} \\ &< e^{-c} \cdot e^c = 1 \end{aligned}$$

which gives the contradiction: $f_\phi(d) < 1$. So, $M(\phi)$ must not be finite.

By Lemma 2, $M(\phi)$ is either finite or co-finite. From above, $M(\phi)$ is not a non-empty finite set. If $t \notin M(\phi)$, then $t \in M(\neg\phi)$. So, we must have $M(\neg\phi)$ is finite. Assume $M(\neg\phi)$ is non-empty, and for some $t_1, \ldots, t_m$, $M(\neg\phi) = \{t_1, \ldots, t_m\}$. Then,

$$f_\phi(c) = 1 - e^{-c} \textstyle\sum_{i=1}^{i=m} \frac{c^{t_i}}{t_i!} < 1 \quad \text{for } c > 0$$

which contradicts $f_\phi(d) = 1$. So, we must have that $M(\neg\phi) = \emptyset$. This gives that f_ϕ is constantly one. So, $\phi \in S_c$ for every c, and thus, $\phi \in S$.

Therefore, $S = \bigcap_c S_c$. □

5 Future Work

The work of [7] and [8] 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 this paper, we fill the "gap" between $p(n) \ll n^{-1}$ and $n^{-1} \ll p(n) \ll n^{-1/2}$ by characterizing the almost sure theories of $U_{n,\frac{c}{n}}$ and giving the form of the function $f_\phi(c)$ for each first-order sentence ϕ. What happens for the gaps at $p(n) = cn^{-1/k}$ for larger k? That is, what does

$$g_{k,\phi} = \lim_{n\to\infty} \Pr[U_{n,cn^{-1/k}} \models \phi]$$

look like for each integer $k > 1$ and first-order sentence ϕ? How does this answer change if ϕ is allowed to be from the extended language of monadic second order logic?

6 Acknowledgments

The author would like to thank Joel Spencer for his many insightful comments on this paper and for his help in clarifying Theorem 1.

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] Herbert B. Enderton. *A Mathematical Introduction to Logic.* Academic Press, San Diego, 1972.

[5] James F. Lynch. Probabilities of sentences about very sparse random graphs. *Random Structures and Algorithms*, 3:33–54, 1992.

[6] Saharon Shelah and Joel H. Spencer. Can you feel the double jump. *Random Structures and Algorithms*, 5(1):191–204, 1994.

[7] Saharon Shelah and Joel H. Spencer. Random sparse unary predicates. *Random Structures and Algorithms*, 5(3):375–394, 1994.

[8] Joel H. Spencer and Katherine St. John. Random unary predicates: Almost sure theories and countable models. Submitted for publication, 1997.

[9] 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.