Distribution of the Length of the Longest Significance Run on a Bernoulli Net and Its Applications

Jihong CHEN and Xiaoming HUO

We consider the length of the longest significance run in a (two-dimensional) Bernoulli net and derive its asymptotic limit distribution. Our results can be considered as generalizations of known theorems in significance runs. We give three types of theoretical results: (1) reliability-style lower and upper bounds, (2) Erdös–Rényi law, and (3) the asymptotic limit distribution. To understand the rate of convergence to the asymptotic distributions, we carry out numerical simulations. The convergence rates in a variety of situations are presented. To understand the relation between the length of the longest significance run(s) and the success probability p, we propose a dynamic programming algorithm to implement *simultaneous* simulations. Insights from numerical studies are important for choosing the values of design parameters in a particular application, which motivates this article. The distribution of the length of the longest significance run in a Bernoulli net is critical in applying a multiscale methodology in image detection and computational vision. Approximation strategies to some critical quantities are discussed.

KEY WORDS: Asymptotic distribution; Bernoulli net; Detection; Dynamic programming algorithm; Erdös–Rényi law.

1. INTRODUCTION

We consider an m-by-n array of nodes—m rows and n columns. Such an array can be considered a grid in a two-dimensional rectangular region, $[1, n] \times [1, m]$. Assume that each node with coordinate (i, j), $1 \le i \le n$, $1 \le j \le m$, is associated with a Bernoulli(p) random variable $X_{i,j}$. If $X_{i,j} = 1$, then the node is called *significant*; otherwise, it is *nonsignificant*. Any two nodes (i_1, j_1) and (i_2, j_2) are *connected* if and only if $|i_1 - i_2| = 1$ and $|j_1 - j_2| \le C$, with C a prescribed positive integer. Define a *chain* of length ℓ as a chain of ℓ *connected* nodes,

$$\{(i_1, j_1), \ldots, (i_1 + \ell - 1, j_\ell) : |j_k - j_{k-1}| \le C, \text{ for } 2 \le k \le \ell\}.$$

A *significance run* refers to a chain with all nodes significant. We call such a system a *Bernoulli net*. Figure 1 illustrates a Bernoulli net and a significance run. We are interested in the length of the longest significance run in this net, which is denoted by L_0.

If L_0 is considered a function of the number of columns, n, then our theoretical results are generalizations of existing results in significance runs (Balakrishnan and Koutras 2002). This becomes more evident as the theorems are described. In fact, our theoretical results are highly parallel to the known results in longest runs.

Our direct motivation is from a statistical detection problem. Arias-Castro, Donoho, and Huo (2003) proposed a method called the *multiscale significance run algorithm* (MSRA) for the detection of curvilinear filaments in noisy images. The main idea is to construct a Bernoulli net. Each node has the value of 1 (significant) or 0 (nonsignificant). Two nodes are defined as "connected" if they are neighbors, that is, they can simultaneously cover a curve of interest. The length of the longest connected significant nodes, called the *longest significance run*, is used as a test statistic. If the length exceeds a certain threshold, then *we conclude that there* exists an embedded curve; otherwise, there is no embedded curve. To formulate this as a well-defined probability problem, we test the *null hypothesis* of a constant success probability p against the *alternative hypothesis* that some nodes, being on a filament with unknown location and length, have a greater probability of success. Under the alternative, L_0 is more likely to exceed (i.e., be greater than) a threshold, which, under the null hypothesis, cannot be exceeded.

Apparently, the longest length (L_0) depends on parameters n, m, p, and C. In the approach of Arias-Castro et al. (2003), the values of these parameters can be chosen. The question is how to choose these parameters so that the power of the test can be maximized. This becomes a design issue. The relation between L_0 and other parameters must be understood. The choice of parameters in the approach of Arias-Castro et al. (2003) is sufficient to guarantee a proof of asymptotic optimality; what we present here is a more precise result. This article does not solve the entire problem, but it is one step in this direction.

This article provides theoretical analysis as well as computational methods for the distribution of L_0 under the null hypothesis. In Section 2 we present product-type upper and lower bounds for the cumulative distribution function of L_0. We also study the asymptotic behavior of the length L_0 as n goes to infinity. We develop computational approaches in Section 3. In Section 3.1 we design an approximation strategy to approximate the true value of the tail probability in the finite-sample case; in Section 3.2 we provide a dynamic programming approach that allows us to study the relation between L_0 and p for a range of p simultaneously. We give detailed proofs for our main theorems in Section 4. In Section 5, we present numerical simulations to illustrate our theoretical results and to evaluate the quality of the suggested approximations. In Section 6, we address the connections between the proposed problem and the methodologies used in image detection and computational vision. Finally, we provide a brief conclusion in Section 7.

Jihong Chen is Graduate Student and Xiaoming Huo is Assistant Professor (E-mail: *xiaoming@isye.gatech.edu*), School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332. This work was supported in part by National Science Foundation grants DMS-01-40587 and DMS-03-46307. The dynamic programming algorithm described in Section 3.2 first came up during a discussion when the second author was visiting David L. Donoho at Stanford University. The comments from the associate editor and four anonymous referees helped improving the content and presentation of this article. In particular, the proof of the lower bound in Theorem 1 was inspired by an anonymous referee, whose comment led to the reference about *association* (Esary, Proschan, and Walkup 1967).

Journal of the American Statistical Association
March 2006, Vol. 101, No. 473, Theory and Methods
DOI 10.1198/016214505000000574

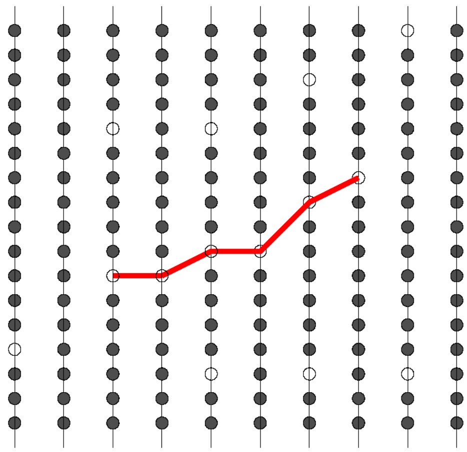

Figure 1. A Significance Run, Where $C = 2$. Hollow nodes are significant.

2. MAIN THEORETICAL RESULTS

Assume that the random variables $X_{i,j}$ are independent. Our first result gives upper and lower bounds for the probability of $P(L_0 < \ell | n, m, C, p)$, where n, m, C, and p are parameters in a Bernoulli net.

Theorem 1. Let $P_\ell = P(L_0 = \ell | \ell, m, C, p)$ denote the probability that the length of the longest run is ℓ, when there are exactly $n = \ell$ columns. We have

$$(1 - P_\ell)^{n-\ell+1} \le P(L_0 < \ell | n, m, C, p) \le [1 - q^m P_\ell]^{n-\ell+1}, \qquad (1)$$

where $q = 1 - p$.

The foregoing is motivated by reliability-focused work (e.g., Papastavridis and Koutras 1993). The techniques used to prove this theorem are purely probabilistic and combinatoric. These bounds are loose, especially when m is large. However, they are useful in the proof of our strong convergence result presented in Theorem 2.

The following lemma introduces a constant, ρ, which is important in the asymptotic distribution of L_0.

Lemma 1. Define $\rho_\ell = P_\ell / P_{\ell-1}$. There exists a constant ρ ($0 < \rho < 1$) that depends only on m, C, and p, and not on n, such that

$$\lim_{\ell \to \infty} \rho_\ell = \rho. \qquad (2)$$

We say a significance run is *across* if and only if it passes all columns. The ratio ρ_ℓ is the conditional probability that there is an across significance run for ℓ columns, conditioning on the fact that there is an across significance run in the previous ($\ell - 1$) columns. We may call this the chance of preserving across significance runs. The foregoing lemma shows that as the number of columns goes to infinity, the chance of preserving across significance runs converges to a constant.

Now we consider an Erdös–Rényi type of result. As mentioned earlier, after fixing the parameters m, C, and p, we can treat L_0 as a function of the number of columns, n. For simplicity, let $L_0(n)$ denote the longest run in such a Bernoulli net.

Theorem 2. As $n \to \infty$,

$$\frac{L_0(n)}{\log_{1/\rho} n} \longrightarrow 1, \quad \text{almost surely.} \qquad (3)$$

This result can be viewed as a generalization of the well-known Erdös–Rényi law (see Petrov 1965; Erdös and Rényi 1970; Erdös and Revesz 1975), which proves that for a one-dimensional sequence ($m = 1$), as $n \to \infty$, (3) holds with ρ replaced by p. Note that when $m = 1$, $P_\ell = p^\ell$ and $\rho_\ell = p$. When $C = 0$, $P_\ell = 1 - (1 - p^\ell)^m$. In both cases, $\rho = \lim_{\ell \to \infty} \rho_\ell = p$. Our result is true for a two-dimensional net, whereas the original Erdös–Rényi law is proved for coin-tossing, which is a one-dimensional sequence.

Using the Chen–Stein Poisson approximation, we prove the following theorem, which gives the asymptotic distribution for $L_0(n)$.

Theorem 3. There exists a constant $A_1 > 0$, that depends only on m, C, and p and not on n, such that for any fixed t, as $n \to \infty$, we have

$$P\big(L_0(n) < \log_{1/\rho} n + t\big) \to \exp\{-A_1 \cdot \rho^t\} \quad \text{as } n \to \infty.$$

The existence of constant A_1 was established in the late part of the proof of this theorem (Sec. 4.4). Because of the nature of the proof, we do not have a specific formula for A_1.

Note that the bounds presented in Theorem 1 are *not* sufficient for deriving this asymptotic distribution. For one-dimensional Bernoulli sequences, there are some similar results. A discussion of this is provided in Section 6.

The foregoing theorems provide a comprehensive description on the asymptotic distribution of the length of the longest significance run, L_0, in a Bernoulli net. The proofs in this article are tailored to the structure of a Bernoulli net. Many techniques are novel and unique to this situation. Proofs of the foregoing theorems are presented in Section 4.

3. MAIN RESULTS IN ALGORITHMIC DEVELOPMENTS

The theoretical results are insightful. However, in the finite-sample case, considering the experimental design task, more numerically specific results must be obtained. We first provide some approximation techniques for the quantity $P(L_0 < \ell | n, m, C, p)$, based on similar quantities designed for smaller regions, for example, $P(L_0 < \ell | i\ell, m, C, p)$, $P(L_0 < \ell | n, j\ell, C, p)$, or $P(L_0 < \ell | i\ell, j\ell, C, p)$, where i and j are positive integers. The main results are presented in (4)–(6) in Section 3.1. Simulations are conducted to test how good these approximations are; these are presented in Section 5.4.

A more ambitious task is to illustrate, for fixed m and n, the connection between L_0 and the value of p. There is a naive approach: For each fixed value of p, run multiple simulations to generate L_0's, then plot the histogram of L_0. This approach is repeated when the value of p is changed.

In Section 3.2 we propose a method that can run simulations for all possible values of p *simultaneously*. Let node (i, j) be associated with a random variable $t_{i,j} \sim \text{Uniform}(0, 1)$. Suppose that we have a realization of the set of random variables $\{t_{i,j} : 1 \le i \le n, 1 \le j \le m\}$. For a probability p, node (i, j) is significant if and only if $t_{i,j} \ge 1 - p$. A realization of a Bernoulli net is given accordingly. One can compute L_0 for this net, which is denoted by $L_0(p)$. A dynamic programming algorithm, described in Section 3.2, shows that the stepwise constant monotone nondecreasing function $L_0(p)$ can be computed for all values $0 \le p \le 1$ from one set of realizations of $\{t_{i,j} : 1 \le i \le n, 1 \le j \le m\}$ *at one time*. Moreover, the time to run of this algorithm is no longer than $mn(n+1)(C+1/2)$, and the space requirement is no more than $mn(n+1)/2$. This is documented in Lemma 2. When the number of values taken for p is large, the proposed simulation approach can save much time, because it does not have to redo simulations for different values of p.

Our numerical approach gives a nice way of illustrating the empirical distribution of the length of the longest run. Figure 2 was computed using the aforementioned method. One easy observation is that for $n = 64$, when $p > .25$, $L_0(p)$ will reach the maximum possible value (which is 64). In other words, one will see a significance run across all of the columns.

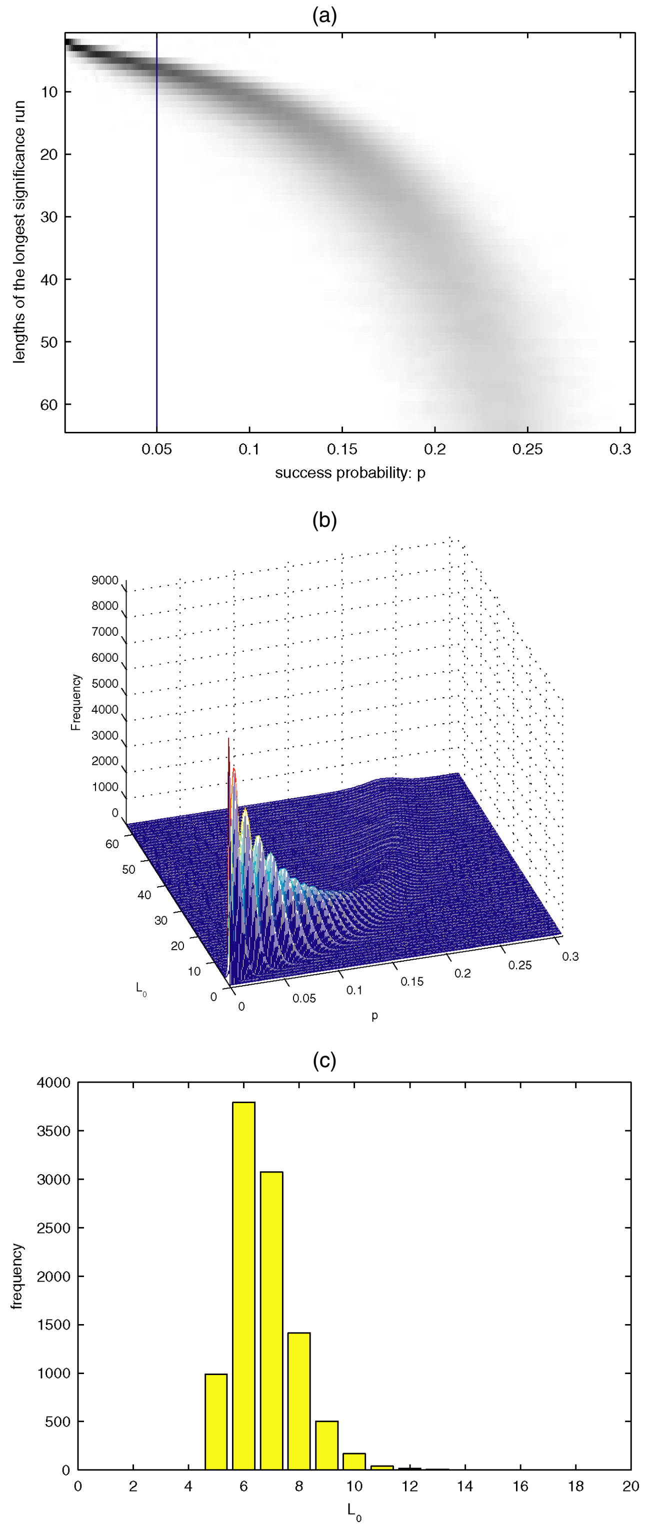

Figure 2. L_0 versus p. (a) An image plot, the distribution of L_0 (under $n = 64$, $m = 128$, $C = 3$) as a function of p ($0 < p < .3075$). The intensity of the image is proportional to the frequency of $L_0(n)$ (which is specified by the y-coordinate) given a value of p (which is the x-coordinate) out of 10,000 simulations. (b) A mesh plot of the same data as in (a). (c) For $p = .05$, the histogram of L_0 based on the same 10,000 simulations. Note this can be viewed as one vertical slice from (a) or, similarly, a slice from (b).

3.1 Numerical Approximation

For large m or n, we derive the following approximations using an approach similar to that of Wallenstein, Naus, and Glaz (1994). We first consider the case when m is large. The longest run in a region $[a, b] \times [c, d]$ represents the longest significance run in the subtable $\{X_{i,j} : a \le i \le b, c \le j \le d\}$. For $1 \le k \le r_1$, let A_k denote the event that the longest run is shorter than ℓ in the subregion $[1, n] \times [(k-1)C\ell + 1, (k-1)C\ell + 2C\ell]$, where we assume that $r_1 = m/(C\ell) - 1$ is an integer. We have

$$\begin{aligned} P(L_0 < \ell | n, m, C, p) &= P\big(A_1 A_2 \cdots A_{r_1}\big) \\ &\approx P(A_1)P(A_2|A_1)P(A_3|A_2)\cdots P\big(A_{r_1}|A_{r_1-1}\big), \end{aligned}$$

where

$$P(A_i|A_{i-1}) = \frac{Q(n, 3C\ell)}{Q(n, 2C\ell)}$$

and $Q(n, iC\ell) = P(L_0 < \ell | n, iC\ell, C, p)$, $i = 2, 3$. Therefore,

$$P(L_0 < \ell | n, m, C, p) \approx Q(n, 2C\ell)\left[\frac{Q(n, 3C\ell)}{Q(n, 2C\ell)}\right]^{m/(C\ell)-2}. \quad (4)$$

Similarly, we can derive an approximation for large n. Let B_k denote the event that the longest run is shorter than ℓ in the subregion $[(k-1)\ell + 1, (k-1)\ell + 2\ell] \times [1, m]$, $1 \le k \le r_2$, where $r_2 = n/\ell - 1$. We have

$$\begin{aligned} P(L_0 < \ell | n, m, C, p) &= P\big(B_1 B_2 \cdots B_{r_2}\big) \\ &\approx P(B_1)P(B_2|B_1)P(B_3|B_2)\cdots P\big(B_{r_2}|B_{r_2-1}\big). \end{aligned}$$

Again,

$$P(L_0 < \ell | n, m, C, p) \approx Q(2\ell, m)\left[\frac{Q(3\ell, m)}{Q(2\ell, m)}\right]^{n/\ell-2}, \quad (5)$$

where $Q(i\ell, m) = P(L_0 < \ell | i\ell, m, C, p)$, $i = 2, 3$.

When both n and m are large, we combine (4) and (5),

$$P(L_0 < \ell | n, m, C, p) \approx Q_2\left(\frac{Q_3}{Q_2}\right)^{n/\ell-2}, \quad (6)$$

where $Q_i = Q_{i2}(Q_{i3}/Q_{i2})^{m/C\ell-2}$ and $Q_{ij} = P(L_0 < \ell|i\ell, jC\ell, C, p)$, $i, j = 2, 3$.

The values of $Q(n, iC\ell)$, $Q(i\ell, m)$, and Q_{ij} mentioned earlier have smaller sample sizes. They can be obtained from simulations by the approach described next.

3.2 A Dynamic Programming Approach to Study the Relation Between L_0 and p

Let $L_0(p)$ denote the length of the longest significance run for a given probability p. We provide a dynamic programming approach to compute $L_0(p)$ for the entire interval [0, 1] for p.

Recall that $t_{i,j} \sim \text{Uniform}(0, 1)$, $1 \le i \le n$, $1 \le j \le m$, are iid uniform random variables. For a given probability p, node (i, j) is significant if $t_{i,j} \ge 1 - p$. Let $L_{i,j}(p)$ denote the length of the longest significance run starting from the leftmost end of the Bernoulli table and ending at node (i, j).

For the nodes in the first column, $(1, j)$, $j = 1, 2, \ldots, m$, we have

$$L_{1,j}(p) = \begin{cases} 0 & \text{if } p < 1 - t_{1,j} \\ 1 & \text{otherwise.} \end{cases}$$

For node (i, j), it is not hard to verify that

$$L_{i,j}(p) = \mathbb{1}\{t_{i,j} \ge 1 - p\} \cdot \left\{1 + \max_{j' \in \Omega(j)} L_{i-1,j'}(p)\right\},$$

where $\Omega(j) = \{j' : |j' - j| \le C, 1 \le j' \le m\}$ denotes the set containing neighboring indices of j. The function $\mathbb{1}\{t_{i,j} > 1 - p\}$ is an indicator function.

When all of the $L_{i,j}(p)$'s are available, the value of $L_0(p)$ satisfies

$$L_0(p) = \max_{i,j} L_{i,j}(p).$$

The function $L_{i,j}(p)$ is piecewise constant and a nondecreasing function of p. Define the break points in the functions $L_{i,j}(p)$ as

$$b_\ell^{(i,j)} = \min\{p : L_{i,j}(p) \ge \ell\}, \qquad \ell = 1, 2, \ldots, n.$$

The value $b_\ell^{(i,j)}$ is the lower bound of the set when the value of $L_{i,j}(\cdot)$ is equal to ℓ. For nodes in the first column, we have

$$b_1^{(1,j)} = 1 - t_{1,j} \qquad \forall j.$$

Because $L_{1,j}(\cdot) \le 1$, we can assume that

$$b_\ell^{(1,j)} = 1 \qquad \forall j \text{ and } \ell > 1.$$

We can derive the updating scheme for the break points as

$$b_1^{(i,j)} = 1 - t_{i,j},$$

and, for $\ell \ge 2$,

$$b_\ell^{(i,j)} = \max\left(b_1^{(i,j)}, \min_{j' \in \Omega(j)} b_{\ell-1}^{(i-1,j')}\right). \tag{7}$$

Note that the foregoing gives a recursive formula with respect to the length of the longest significance run ℓ that ends at this node and the column index i. Define

$$b_\ell^* = \min\{p : L_0(p) \ge \ell\}.$$

It is not hard to see that

$$b_\ell^* = \min_{i,j} b_\ell^{(i,j)} \qquad \forall \ell = 1, 2, \ldots, n.$$

The foregoing gives an algorithm for computing $L_0(p)$ for the entire interval $0 \le p \le 1$.

We now consider the time and space requirements of the proposed algorithm. For a node at column i, there are at most i break points, because the maximum length of a significance run up to this node is i. For each break point, according to (7), there are at most $(2C + 1)$ previous break points to compare. Hence, it takes at most $m(2C + 1)i$ operations to compute break points for all of the nodes in column i. The run time of the entire algorithm is at most

$$\sum_{i=1}^{n} m(2C + 1)i = \frac{m(2C + 1)n(n + 1)}{2}.$$

Obviously, the space is no more than

$$\sum_{i=1}^{n} mi = \frac{mn(n + 1)}{2}.$$

We have proved the following lemma.

Lemma 2. Given a realization of variables $\{t_{i,j} : 1 \le i \le n, 1 \le j \le m\}$, there is a dynamic programming algorithm for computing the value of function $L_0(p)$ for all values $0 \le p \le 1$ simultaneously. The computational time is upper-bounded by $mn(n + 1)(C + 1/2)$, and the required space is no more than $mn(n + 1)/2$.

The foregoing can be used in carrying out simultaneous simulations. For each simulation, the results of break points (i.e., b_ℓ^*'s, $\ell = 1, 2, \ldots, n$) are arranged in a $1 \times n$ vector. By conducting N (e.g., $N = 10{,}000$ as in Fig. 2) simulations, we obtain a matrix of size $N \times n$. Given p, the probability $P(L_0(p) \ge \ell)$ is estimated by the fraction of b_ℓ^*'s that are smaller than p. Figure 2 is generated in this way. Note that Figure 2 gives a nice illustration of the relation between $L_0(p)$ and p.

4. PROOFS OF THE THEOREMS

Recall that $P(L_0 \ge \ell|n, m, C, p)$ is the probability of $L_0 \ge \ell$ in an $m \times n$ Bernoulli net with a common success probability p. Obviously, for a simple Bernoulli net with $C = 0$, we have

$$\begin{aligned} P(L_0 \ge \ell|n, m, 0, p) &= 1 - P(L_0 < \ell|n, m, 0, p) \\ &= 1 - \left[1 - P(L_0 \ge \ell|n, 1, 0, p)\right]^m, \end{aligned}$$

where $P(L_0 \ge \ell|n, 1, 0, p)$ is the distribution of the longest run in a one-dimensional Bernoulli sequence.

In what follows, some of the approaches and techniques for handling one-dimensional longest run are generalized to handle the two-dimensional problem.

4.1 Proof of Theorem 1

For the one-dimensional case ($m = 1$), the following simple bound was originally developed in a reliability-focused work (Papastavridis and Koutras 1993):

$$(1 - p^\ell)^{n-\ell+1} \le P(L_0 < \ell|n, 1, 0, p) \le (1 - qp^\ell)^{n-\ell+1},$$

where $q = 1 - p$. An extension of the foregoing bounds to a two-dimensional Bernoulli net yields Theorem 1.

To prove the theorem, we introduce the following notation:

- E_i, the event that the longest run is shorter than ℓ in the subregion $[i, i+\ell-1] \times [1, m]$, $1 \le i \le n-\ell+1$. That is, the longest run among nodes $\{X_{a,b} : i \le a \le i+\ell-1, 1 \le b \le m\}$ is shorter than ℓ. The following statements can be interpreted in the same way
- F_i, the event that the longest run is shorter than ℓ in the subregion $[1, i] \times [1, m]$, $\ell \le i \le n$
- A', the complement of the set A
- G_i, the event that there is no significant node on the $(i-\ell)$th column.

4.1.1 Upper Bound. For the upper bound, we have

$$\begin{aligned} P(L_0 < \ell | n, m, C, p) &= P(F_n) \\ &= P(F_\ell) \prod_{i=\ell+1}^{n} \frac{P(F_i)}{P(F_{i-1})} \\ &= P(F_\ell) \prod_{i=\ell+1}^{n} \left[1 - P(F_i' | F_{i-1})\right]. \end{aligned}$$

The foregoing are simply basic probability derivations. To prove the upper bound, we need only to verify that

$$\begin{aligned} P(F_i' | F_{i-1}) &\ge P(F_i' | G_i F_{i-1}) \cdot P(G_i | F_{i-1}) \\ &\ge P_\ell \cdot P(G_i) \\ &= (1-p)^m P_\ell. \end{aligned}$$

The first inequality is obvious. By definition, we can easily see that $P(F_i' | G_i F_{i-1}) = P_\ell$. To make the second inequality hold, we need

$$P(G_i | F_{i-1}) \ge P(G_i). \tag{8}$$

The foregoing can be seen from

$$P(F_{i-1}) \le P(F_{i-\ell-1}) = P(F_{i-1} | G_i), \tag{9}$$

where the first inequality is from the definition of the F_i's and the second equality is straightforward. The foregoing two inequalities (8) and (9) are equivalent. Hence we have proved the upper bound.

4.1.2 Lower Bound. For the lower bound, we need to prove that

$$P(E_1 \cap E_2 \cap \cdots \cap E_{n-\ell+1}) \ge \prod_{i=1}^{n-\ell+1} P(E_i). \tag{10}$$

The properties of *association* among random variables, described by Esary et al. (1967), will be used. Recall that n random variables $T_1, T_2, \ldots, T_n$ are associated if $\mathrm{cov}[f(\mathbf{T}), g(\mathbf{T})] \ge 0$, for all nondecreasing functions f and g, for which the expectations $\mathbf{E}(f)$, $\mathbf{E}(g)$, and $\mathbf{E}(fg)$ exist. It is known that

- Nondecreasing functions of associated random variables are associated [Esary et al. 1967, (P4)] and
- Independent random variables are associated (Esary et al. 1967, thm. 2.1).

Recall that the random variables $X_{i,j}$, $1 \le i \le n$, $1 \le j \le m$, are independent, and hence they are associated. Consider a new set of random variables, $D_i = \mathbb{1}\{E_i'\}$, $i = 1, 2, \ldots, n-\ell+1$. Evidently, $\mathbb{1}\{E_i'\}$ is a nondecreasing function of random variables $X_{a,b}$, then $i \le a \le i+\ell-1$, $1 \le b \le m$. Hence random variables $D_1, D_2, \ldots, D_{n-\ell+1}$ are associated. According to Esary et al. (1967, thm. 4.1), we have

$$\begin{aligned} &P(D_1 = 0, D_2 = 0, \ldots, D_{n-\ell+1} = 0) \\ &\qquad \ge P(D_1 = 0) \cdot P(D_2 = 0) \cdots P(D_{n-\ell+1} = 0). \end{aligned}$$

It is not hard to verify that $P(D_1 = 0, D_2 = 0, \ldots, D_{n-\ell+1} = 0) = P(E_1 \cap E_2 \cap \cdots \cap E_{n-\ell+1})$ and $P(D_i = 0) = P(E_i)$, for $i = 1, 2, \ldots, n-\ell+1$. Hence we have proved (10).

From all of the foregoing, we have the following result on the lower bound:

$$\begin{aligned} P(L_0 < \ell | n, m, C, p) &= P(E_1 \cap E_2 \cap \cdots \cap E_{n-\ell+1}) \\ &\ge \prod_{i=1}^{n-\ell+1} P(E_i) \\ &= (1 - P_\ell)^{n-\ell+1}. \end{aligned}$$

This proves Theorem 1.

A one-dimensional version of this result (i.e., when $m = 1$) was given in by Muselli (1997), whose lemma 1 gave a pure (and interesting) combinatoric proof. More advanced results in one-dimensional situation have been given by Muselli (2000). The application of association in this problem seems to greatly simplify the proof.

4.2 Proof of Lemma 1

Without loss of generality, we need only to consider $C \ge 1$. The case of $C = 0$ is trivial and has been mentioned in Section 2.

Let $\{\mathbf{x}_i, i = 1, 2, \ldots\}$ be a Markov chain, where $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,m})$ is a vector of length m. $x_{i,j}$ denotes the state of node (i, j). We have $x_{ij} = 1$ if the length of the longest run starting from the left end and ending at node (i, j) is equal to i; otherwise, $x_{ij} = 0$. Let S be the set of all possible values of $\mathbf{x}_i$. The cardinality of S is 2^m.

For $j = 1, \ldots, m$, let $\Omega(j) = \{j' : |j' - j| \le C, 1 \le j' \le m\}$ be the set containing neighboring indices of j. For two states $\mathbf{s}_1, \mathbf{s}_2 \in S$, the transition probability is

$$P_{\mathbf{s}_1 \mathbf{s}_2} = p^a (1-p)^b \mathbb{1}\{c = 0\}, \tag{11}$$

where

$$a = \sum_{j=1}^{m} \mathbb{1}\Big\{ \max_{j' \in \Omega(j)} \mathbf{s}_1(j') = 1, \mathbf{s}_2(j) = 1 \Big\},$$

$$b = \sum_{j} \mathbb{1}\Big\{ \max_{j' \in \Omega(j)} \mathbf{s}_1(j') = 1, \mathbf{s}_2(j) = 0 \Big\},$$

and

$$c = \sum_{j} \mathbb{1}\Big\{ \max_{j' \in \Omega(j)} \mathbf{s}_1(j') = 0,\ \mathbf{s}_2(j) = 1 \Big\}.$$

Here $\mathbf{s}_1(j')$ and $\mathbf{s}_2(j)$ denote the values of states $\mathbf{s}_1$ and $\mathbf{s}_2$ at the j'th and jth rows. Obviously, $c = 0$ when $\mathbf{s}_1 = \mathbf{s}_2$. Therefore, $P_{\mathbf{ss}} > 0\ \forall \mathbf{s} \in S$.

Letting $\pi_{\mathbf{s}}^i = P\{\mathbf{x}_i = \mathbf{s} | \mathbf{x}_i \neq \mathbf{0}\}$, we have

$$\begin{aligned}\rho_\ell &= \frac{P_\ell}{P_{\ell-1}} = \frac{P\{\mathbf{x}_\ell \neq \mathbf{0}\}}{P\{\mathbf{x}_{\ell-1} \neq \mathbf{0}\}} \\ &= P\{\mathbf{x}_\ell \neq \mathbf{0} | \mathbf{x}_{\ell-1} \neq \mathbf{0}\} \\ &= \sum_{\mathbf{s}\neq\mathbf{0}} \pi_{\mathbf{s}}^{\ell-1}(1 - P_{\mathbf{s0}}).\end{aligned}$$

Now define another Markov chain, $\mathbf{y}_i = \{\mathbf{x}_i | \mathbf{x}_i \neq \mathbf{0}\}$, with state space $S\backslash\{\mathbf{0}\}$. The transition probability of the new Markov chain is

$$P'_{\mathbf{s}_1\mathbf{s}_2} = \frac{P_{\mathbf{s}_1\mathbf{s}_2}}{\sum_{\mathbf{s}\neq\mathbf{0}} P_{\mathbf{s}_1\mathbf{s}}}, \qquad \mathbf{s}_1, \mathbf{s}_2 \in S\backslash\{\mathbf{0}\}. \tag{12}$$

It is obvious that $P'_{\mathbf{ss}} > P_{\mathbf{ss}} > 0$. Therefore, the Markov chain is aperiodic.

Moreover, it is easy to see that any other state is accessible from the state $(1, 1, \dots, 1)$ in one step. Also, the state $(1, 1, \dots, 1)$ is accessible from any other state in m steps, so all of the states communicate with each other. Therefore, the new Markov chain is irreducible.

Because the Markov chain $\{\mathbf{y}_i, i = 1, 2, \dots\}$ is finite, aperiodic, and irreducible, there exists limiting distribution $\pi_{\mathbf{s}}$ such that (Kulkarni 1995)

$$\lim_{\ell\to\infty} \pi_{\mathbf{s}}^\ell = \pi_{\mathbf{s}}.$$

Therefore,

$$\lim_{\ell\to\infty} \rho_\ell = \sum_{\mathbf{s}\neq\mathbf{0}} \pi_{\mathbf{s}} \cdot (1 - P_{\mathbf{s0}}) = \rho. \tag{13}$$

4.3 Proof of Theorem 2

For any real value x, we define

$$L_0(x) = L_0(\lfloor x \rfloor).$$

First, we prove that

$$\frac{L_0(e^k)}{\log_{1/\rho} e^k} \longrightarrow 1 \quad \text{as } k \to \infty \text{ almost surely.} \tag{14}$$

To prove (14), we need to prove that, $\forall\, \epsilon > 0$,

$$P\left(\left|\frac{L_0(e^k)}{\log_{1/\rho} e^k} - 1\right| > \epsilon \text{ infinitely often}\right) = 0.$$

According to the Borel–Cantelli lemmas, it is sufficient to prove that

$$\sum_k P\left(\left|\frac{L_0(e^k)}{\log_{1/\rho} e^k} - 1\right| > \epsilon\right) < \infty. \tag{15}$$

From (2), $\forall\, \delta > 0, \exists\, \ell_0$ such that when $\ell > \ell_0$,

$$\rho^{1+\delta} \le \frac{P_\ell}{P_{\ell-1}} \le \rho^{1-\delta}.$$

Therefore,

$$P_{\ell_0}\rho^{(1+\delta)(\ell-\ell_0)} \le P_\ell \le P_{\ell_0}\rho^{(1-\delta)(\ell-\ell_0)}.$$

From (1), we have

$$\begin{aligned}\left[1 - a_1\rho^{\ell(1-\delta)}\right]^{\lfloor e^k \rfloor - \ell + 1} &\le P\big(L_0(e^k) < \ell\big) \\ &\le \left[1 - a_2\rho^{\ell(1+\delta)}\right]^{\lfloor e^k \rfloor - \ell + 1},\end{aligned}$$

where $a_1 = P_{\ell_0}\rho^{-\ell_0(1-\delta)}$, $a_2 = (1-p)^m P_{\ell_0}\rho^{-\ell_0(1+\delta)}$. We have

$$\begin{aligned}&P\left(\left|\frac{L_0(e^k)}{\log_{1/\rho} e^k} - 1\right| > \epsilon\right) \\ &\quad = P\big(L_0(e^k) > (1+\epsilon)\log_{1/\rho} e^k\big) \\ &\qquad + P\big(L_0(e^k) < (1-\epsilon)\log_{1/\rho} e^k\big) \\ &\quad \le 1 - \left\{1 - a_1\rho^{(1-\delta)\lfloor(1+\epsilon)\log_{1/\rho} e^k\rfloor}\right\}^{\lfloor e^k\rfloor - \lfloor(1+\epsilon)\log_{1/\rho} e^k\rfloor + 1} \\ &\qquad + \left\{1 - a_2\rho^{(1+\delta)\lceil(1-\epsilon)\log_{1/\rho} e^k\rceil}\right\}^{\lfloor e^k\rfloor - \lceil(1-\epsilon)\log_{1/\rho} e^k\rceil + 1} \\ &\quad \le 1 - \left\{1 - a_1\rho^{(1-\delta)[(1+\epsilon)\log_{1/\rho} e^k - 1]}\right\}^{e^k - (1+\epsilon)\log_{1/\rho} e^k + 2} \\ &\qquad + \left\{1 - a_2\rho^{(1+\delta)[(1-\epsilon)\log_{1/\rho} e^k + 1]}\right\}^{e^k - (1-\epsilon)\log_{1/\rho} e^k - 1}.\end{aligned} \tag{16}$$

There exists k_0 such that when $k > k_0$,

$$(1+\epsilon)\log_{1/\rho} e^k \ge 2 \quad \text{and} \quad (1-\epsilon)\log_{1/\rho} e^k + 1 \le \frac{e^k}{2}. \tag{17}$$

Choose $\delta = \frac{1}{4}\epsilon$; then

$$(1-\delta)(1+\epsilon) \ge 1 + \frac{\epsilon}{2} \quad \text{and} \quad (1+\delta)(1-\epsilon) \le 1 - \frac{\epsilon}{2}. \tag{18}$$

Substituting (17) and (18) into (16), we have

$$\begin{aligned}&P\left(\left|\frac{L_0(e^k)}{\log_{1/\rho} e^k} - 1\right| > \epsilon\right) \\ &\qquad \le 1 - \left[1 - a'_1 e^{-(1+\epsilon/2)k}\right]^{e^k} + \left[1 - a'_2 e^{-(1-\epsilon/2)k}\right]^{e^k/2},\end{aligned}$$

where $a'_1 = a_1\rho^{-(1-\delta)}$ and $a'_2 = a_2\rho^{(1+\delta)}$. Considering

$$\lim_{k\to\infty}\left[1 - a'_1 e^{-(1+\epsilon/2)k}\right]^{e^{(1+\epsilon/2)k}/a'_1} = e^{-1}$$

for $\delta_1 = e^{-1}/2$, there exists k_1 such that when $k > k_1$,

$$\left[1 - a'_1 e^{-(1+\epsilon/2)k}\right]^{e^{(1+\epsilon/2)k}/a'_1} \ge e^{-1} - \delta_1$$

and

$$\left[1 - a'_2 e^{-(1-\epsilon/2)k}\right]^{e^{(1-\epsilon/2)k}/a'_2} \le e^{-1} + \delta_1.$$

Let $b_1 = e^{-1}/2$, and $b_2 = 3e^{-1}/2$. Then for $k > k_2 = \max(k_0, k_1)$, we have

$$\begin{aligned}&P\left(\left|\frac{L_0(e^k)}{\log_{1/\rho} e^k} - 1\right| > \epsilon\right) \\ &\quad \le 1 - b_1^{a'_1 e^{-k\epsilon/2}} + b_2^{(a'_2/2)e^{k\epsilon/2}} \\ &\quad \le \left(a'_1 \ln\frac{1}{b_1}\right)e^{-k\epsilon/2} + b_2^{(a'_2/4)k\epsilon}.\end{aligned} \tag{19}$$

Note that $1 - b_1^x \le -x\ln b_1$ and $e^x \ge x$. We have

$$\sum_{k=1}^{\infty} P\left(\left|\frac{L_0(e^k)}{\log_{1/\rho} e^k} - 1\right| > \epsilon\right) < \infty.$$

Therefore, (14) holds as $k \to \infty$.

Denote

$$f_k = \frac{L_0(e^k)}{\log_{1/\rho} e^k}.$$

For any n, there exists k_n such that $e^{k_n} \le n \le e^{k_n+1}$. Because both $L_0(x)$ and $\log_{1/\rho} x$ are increasing functions, we have

$$L_0(e^{k_n}) \le L_0(n) \le L_0(e^{k_n+1})$$

and

$$\frac{1}{\log_{1/\rho} e^{k_n+1}} \le \frac{1}{\log_{1/\rho} n} \le \frac{1}{\log_{1/\rho} e^{k_n}},$$

which, consequently, lead to (by multiplying the foregoing)

$$\frac{L_0(e^{k_n})}{\log_{1/\rho} e^{k_n+1}} \le \frac{L_0(n)}{\log_{1/\rho} n} \le \frac{L_0(e^{k_n+1})}{\log_{1/\rho} e^{k_n}},$$

which is equivalent to

$$\frac{L_0(e^{k_n})}{\log_{1/\rho} e^{k_n}} \frac{\log_{1/\rho} e^{k_n}}{\log_{1/\rho} e^{k_n+1}} \le \frac{L_0(n)}{\log_{1/\rho} n} \le \frac{L_0(e^{k_n+1})}{\log_{1/\rho} e^{k_n+1}} \frac{\log_{1/\rho} e^{k_n+1}}{\log_{1/\rho} e^{k_n}}.$$

Therefore,

$$f_{k_n} \frac{k_n}{k_n+1} \le \frac{L_0(n)}{\log_{1/\rho} n} \le f_{k_n+1} \frac{k_n+1}{k_n}.$$

We have that

$$\frac{L_0(n)}{\log_{1/\rho} n} \to 1 \quad \text{when } n \to \infty.$$

From all of the foregoing, Theorem 2 is proved.

4.4 Proof of Theorem 3

The key idea is to apply the Chen–Stein Poisson approximation, which was described by Arratia, Gordon, and Waterman (1990, sec. 3). Recall that E_i was defined in Section 4.1 as the event that there is no length ℓ significance run in the region $[i, i+\ell-1] \times [1, m]$, $1 \le i \le n-\ell+1$. Define random variable $Y_i = \mathbb{1}\{E_i'\}$. We have $Y_i = 1$ if and only if there is a length ℓ significance run between the ith column and the $(i+\ell-1)$st column and $Y_i = 0$ otherwise. (Recall that D_i was defined the same way as in a previous proof. We prefer to change the notation because of a different use of these random variables.) Define $Z_i = Y_i \prod_{k=i-\ell+1}^{i-1} (1 - Y_k)$. For notational simplicity, we assume $Y_k = 0$ if $k \le 0$. The Z_i is an indicator function of whether or not there is a *clump* starting at the ith column. Here "clump" is a concept used in the Poisson approximation (see Arratia et al. 1990). A clump is made by ℓ consecutive columns containing an across-significance run, conditioning on no previous clumps overlapping with the present one. Let $W = \sum_{i=1}^{n-\ell+1} Z_i$. The number of clumps is equal to W. The main idea of the Poisson approximation is that the distribution of the random variable W can be approximated by Poisson(λ), where the Poisson parameter λ can be computed directly. Details follow.

To verify the conditions for the Poisson approximation, we define the neighborhood of α, $1 \le \alpha \le n$, as $B_\alpha = \{\beta : |\alpha - \beta| < 2\ell, 1 \le \beta \le n\}$. If $\beta \notin B_\alpha$, then clearly we have $|\alpha - \beta| \ge 2\ell$. Now the random variable Z_α depends only on columns from $\alpha - \ell + 1$ to $\alpha + \ell - 1$. Similarly, random variable Z_β depends only on columns from $\beta - \ell + 1$ to $\beta + \ell - 1$. Thus if $\beta \notin B_\alpha$, then Z_α and Z_β are independent. It follows that $b_3 = 0$. (Note that constants b_1, b_2, and b_3 are as defined in Arratia et al. 1990, sec. 3.) For b_1, we have

$$\begin{aligned} b_1 &= \sum_{\alpha=1}^{n} \sum_{\beta \in B_\alpha} E(Z_\alpha) E(Z_\beta) \\ &= \sum_{\alpha=1}^{n} \sum_{\beta \in B_\alpha} P(Z_\alpha = 1) P(Z_\beta = 1) \\ &< \sum_{\alpha=1}^{n} P(Y_\alpha = 1) \sum_{\beta \in B_\alpha} P(Y_\beta = 1) \\ &< n \cdot 4\ell P_\ell^2, \end{aligned}$$

where P_ℓ is the constant defined in Theorem 1.

For b_2, we have

$$\begin{aligned} b_2 &= \sum_{\alpha=1}^{n} \sum_{\beta \in B_\alpha, \beta \ne \alpha} E(Z_\alpha Z_\beta) \\ &= 2 \sum_{\alpha=1}^{n} \sum_{\beta \in B_\alpha, \beta > \alpha} E(Z_\alpha Z_\beta) \\ &= 2 \sum_{\alpha=1}^{n} \sum_{\beta \in B_\alpha, \beta > \alpha} P(Z_\alpha = 1 \text{ and } Z_\beta = 1) \\ &= 2 \sum_{\alpha=1}^{n} P(Z_\alpha = 1) \cdot \sum_{\beta \in B_\alpha, \beta > \alpha} P(Z_\beta = 1 | Z_\alpha = 1) \\ &< 2 \sum_{\alpha=1}^{n} P(Y_\alpha = 1) \cdot \sum_{\alpha+\ell \le \beta < \alpha + 2\ell} P(Y_\beta = 1 | Z_\alpha = 1) \\ &= 2 \sum_{\alpha=1}^{n-\ell+1} P_\ell \sum_{\alpha+\ell \le \beta < \alpha + 2\ell} P(Y_\beta = 1) \\ &\le 2n\ell P_\ell^2. \end{aligned}$$

In the foregoing, we used the following results:

- $P_\ell = P(Y_\alpha = 1)$, for $1 \le \alpha \le n - \ell + 1$, according to the definition of P_ℓ.
- When $\alpha + \ell \le \beta$, Y_β and Z_α are independent, and we have $P(Y_\beta = 1 | Z_\alpha = 1) = P(Y_\beta = 1)$.
- If $\alpha > n - \ell + 1$, then we have $P(Y_\alpha = 1) = 0$.

Now we consider the Poisson parameter λ. Recall that $\lambda = E(W)$. It is easy to see that

$$\lambda \approx nE(Z_\ell) \quad \text{as } n \to \infty. \tag{20}$$

For the expectation $E(Z_\ell)$, we can easily verify the following bounds:

1. $$E(Z_\ell) \le P(Y_\ell = 1) = P_\ell.$$

2. $$\begin{aligned} E(Z_\ell) &= P(Y_\ell = 1, Y_{\ell-1} = \cdots = Y_1 = 0) \\ &> P(E'_\ell \cap G_{2\ell-1}) \\ &= q^m P_\ell, \end{aligned}$$

   where $G_{2\ell-1}$ and E_i are as defined in Section 4.1 and $q = 1 - p$.

Based on the foregoing, there is a constant A_2, $q^m \le A_2 \le 1$, which depends only on m, C, and p and not on n, and we have $E(Z_\ell) = A_2 P_\ell$.

From the foregoing, for an arbitrarily small $\delta_1 > 0$, when n is sufficiently large, we have

$$1 - \delta_1 \le \frac{\lambda}{nA_2 P_\ell} \le 1 + \delta_1.$$

From Lemma 1, there exists a constant $A_3 > 0$ for an arbitrarily small constant $\delta_2 > 0$, and when n is sufficiently large, we have

$$1 - \delta_2 \le \frac{P_\ell}{A_3 \rho^\ell} \le 1 + \delta_2.$$

Define $A_1 = A_2 A_3$. From the foregoing, we have, for an arbitrarily small constant $\delta_3 > 0$,

$$1 - \delta_3 \le \frac{\lambda}{nA_1 \rho^\ell} \le 1 + \delta_3.$$

In fact, $\delta_3 = \delta_1 + \delta_2 + \delta_1\delta_2$.

Recall that the Poisson approximation gives (Arratia et al. 1990, lemma 2)

$$|P(W = 0) - e^{-\lambda}| \le \min(1, \lambda^{-1})(b_1 + b_2 + b_3).$$

Hence we have

$$\begin{aligned} \left|P(L_0(n) < \ell) - e^{-nA_1\rho^\ell}\right| &\le \min\left(1, \frac{1}{nA_1\rho^\ell}\right) 6n\ell P_\ell^2 \\ &\le \frac{1}{A_2} 6\ell P_\ell (1 + \delta_2) \\ &\le 6\ell \frac{A_3}{A_2} \rho^\ell (1 + \delta_2)^2. \end{aligned}$$

Now let $\ell = \log_{1/\rho} n + t$. One can easily observe that $\ell\rho^\ell \to 0$ and $A_1 n \rho^\ell = A_1 \rho^t$. Hence Theorem 3 is proved.

5. SIMULATIONS

In this section we present several numerical examples to illustrate our theoretical results. We also present numerical comparisons of simulated distributions with their approximations.

5.1 Value of ρ

Table 1 gives the exact values of ρ for different p's and m's: $m = 4, 8, 10$. The Markov chain approach that was described in the proof of Theorem 1 is used. We write the matrix $\mathbf{P} = \{P_{s_1 s_2}\}$, which was defined by (12). The limiting distribution $\pi = \{\pi_s\}$ is computed by solving the system of linear equations $\pi = \pi\mathbf{P}$. The value of ρ can be obtained from (13). It is not hard to show that the algorithmic complexity is $O(2^{3m})$.

Table 1. The Values of ρ for Different Values of m and p, When $C = 1$

	p					
m	.1	.2	.3	.4	.5	.6
4	.2444	.4564	.6341	.7758	.8804	.9482
8	.2654	.4955	.6869	.8363	.9383	.9876
10	.2691	.5022	.6958	.8467	.9486	.9930

5.2 Empirical Distributions of $L_0(n)$

Figure 3 shows the simulated distributions of $L_0(n)$ for $n = 16, 32, 64, 128$ when $m = 64$, $C = 1$, and $p = .2$. The distribution curves are highly skewed to the right, and the expectation, $E[L_0(n)]$, moves toward ∞ approximately at the rate $\log_{1/\rho} n$. Doubling the value of n makes the expectation $E[L_0(n)]$ shifting to the right by a constant. Note that in the simulations, $E[L_0(n)]$ was approximated by the sample average of the simulated $L_0(n)$'s. Figure 4 shows the distribution of $L_0(n)$ for different values of m and C.

5.3 Convergence Rate to the Erdös–Rényi Law

We also study the convergence rate of (3). Fixing $m = 8$ and $C = 1$ for $n = 10, 20, \ldots, 100$, Figure 5(a) plots the function (as a function of n)

$$\frac{\hat{E}[L_0(n)]}{\log_{1/\rho}(n)},$$

where $\hat{E}[L_0(n)]$ denotes the sample average of $L_0(n)$ from 10,000 simulations. In Figure 5(b), the foregoing sample average ($\hat{E}[L_0(n)]$) is replaced by the sample median. The fluctuation in the latter case is due mainly to the granularity of the $L_0(n)$; note that the median of the $L_0(n)$ can take only integral values.

5.4 Approximation Formulas

Next we compare simulated probabilities $P(L_0 \ge \ell | n, m, C, p)$ with the approximations based on (4), (5), and (6). In

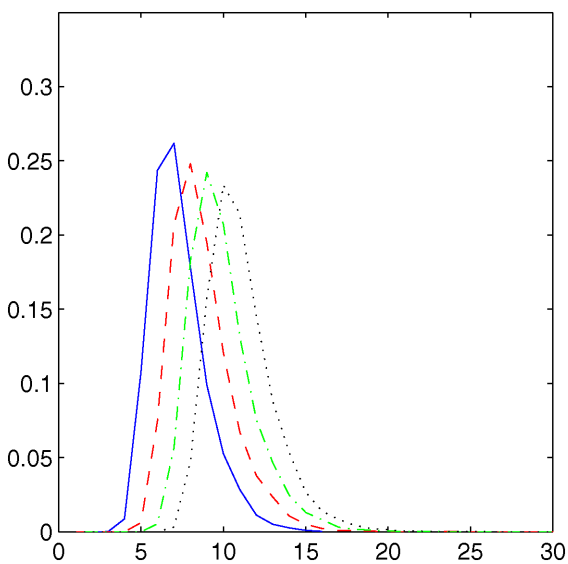

Figure 3. Simulated Distributions of $L_0(n)$ With $m = 64$, $C = 1$, and $p = .2$ (——, $n = 16$; – –, $n = 32$; · – ·, $n = 64$; · · · ·, $n = 128$). The number of simulations is 10,000. The horizontal axis contains the values of $L_0(n)$. The vertical axis contains the sample proportions.

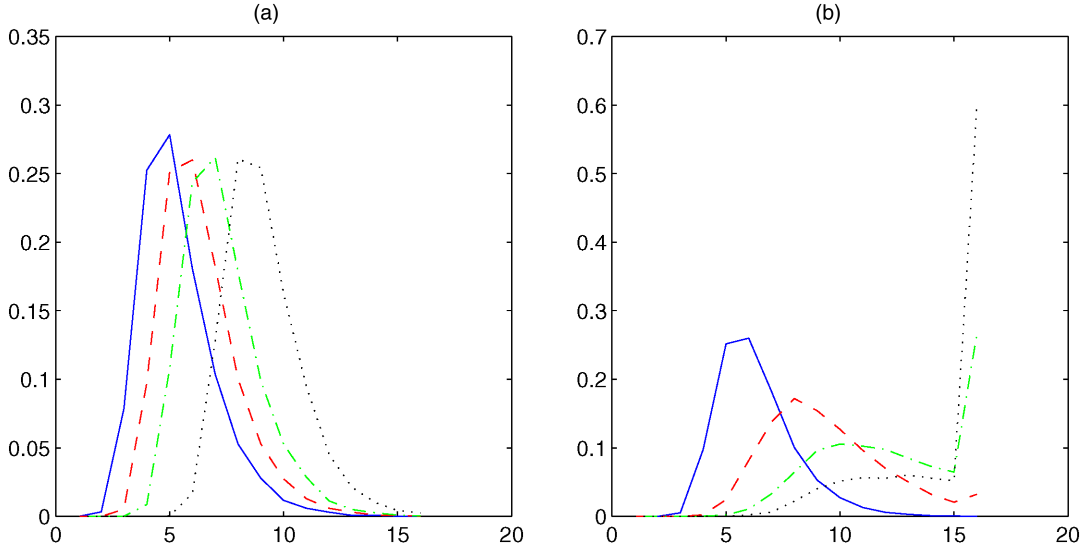

Figure 4. Distributions of L_0 for (a) Different m's With $n = 16$, $C = 1$, and $p = .2$ and (b) Different C's With $n = 128$, $m = 64$, and $p = .2$. In each plot the horizontal axis contains the values of $L_0(n)$, and the vertical axis contains the sample proportions [(a) ——, m = 16; – –, m = 32; ·–·, m = 64; ·····, m = 128; (b) ——, c = 1; – –, c = 2; ·–·, c = 3; ·····, c = 4].

Tables 2–4, "S" stands for the simulated probabilities and "A" stands for the approximated probabilities. In Table 2 the approximated $P(L_0 \geq \ell | 16, m, 1, p)$ requires two simulated probabilities, $P(L_0 \geq 8 | 16, 2\ell, 1, p)$ and $P(L_0 \geq 8 | 16, 3\ell, 1, p)$. Similarly, in Table 3 the approximated $P(L_0 \geq \ell | n, 16, 1, p)$ also requires two simulated probabilities, $P(L_0 \geq \ell | 2\ell, 16, 1, p)$, and $P(L_0 \geq \ell | 3\ell, 16, 1, p)$. In Table 4 the approximated $P(L_0 \geq 8 | n, m, 1, p)$ requires four simulated probabilities, $P(L_0 \geq 8 | 16, 16, 1, p)$, $P(L_0 \geq 8 | 24, 16, 1, p)$, $P(L_0 \geq 8 | 16, 24, 1, p)$, and $P(L_0 \geq 8 | 24, 24, 1, p)$. In all of the foregoing cases, we have $C = 1$, and we allow p to vary. We observe that the approximated probabilities are close to the simulated probabilities.

6. RELATED WORKS AND DISCUSSION

6.1 Our Motivation

As mentioned earlier, our major motivation is from an image detection project. Figure 6 gives such an illustration. For computational details we refer to work of Huo, Chen, and Donoho (2003), which is also downloadable from the second author's publication website, *http://isye.gatech.edu/~xiaoming/publication/*. Here we provide a brief summary of the essence of the method.

Consider tilted rectangles, as shown in Figures 6(b) and 6(d). They are called *axoids* (Huo et al. 2003), which are multiscale objects with different widths and heights, taking different orientations. They are a part of a multiscale methodology developed

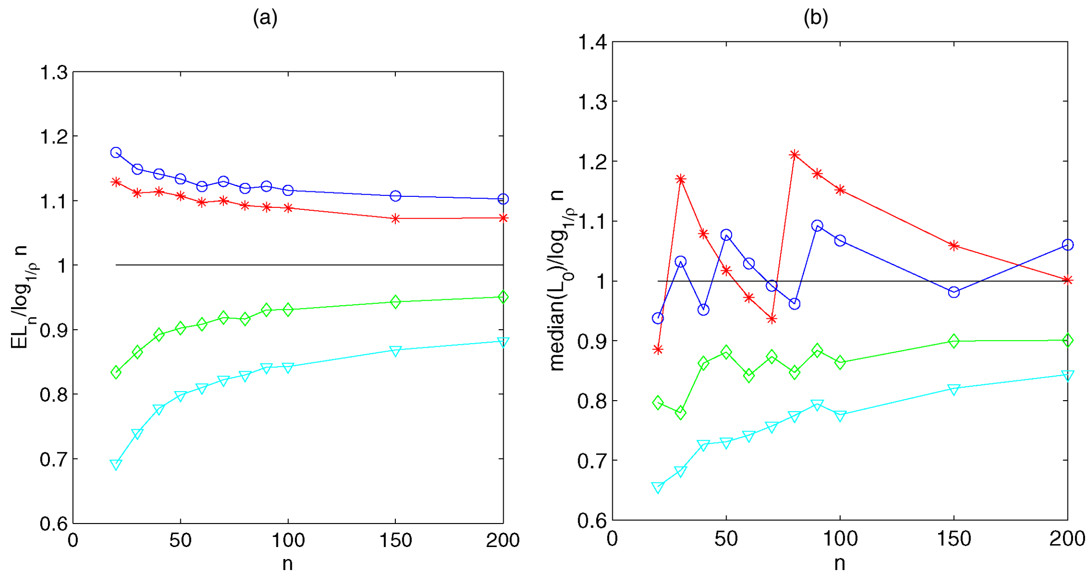

Figure 5. Plots of (a) $\hat{E}[L_0(n)]/\log_{1/\rho}(n)$ Against n for a Range of p's and (b) Median$(L_0(n))/\log_{1/\rho}(n)$ Against n for a Range of p's (–✳–, p = .05; –○–, p = .1; –◇–, p = .35; –▽–, p = .4).

Table 2. Comparisons of the Simulated $P(L_0 \geq \ell \mid n = 16, m, C = 1, p)$ With Approximations by (4) When $m = 32, 64, 128$ and $\ell = 4, 8$

	$\ell = 4$				$\ell = 8$			
	$p = .05$		$p = .10$		$p = .20$		$p = .25$	
m	S	A	S	A	S	A	S	A
32	.0451	.0512	.4294	.4310	.2039	.2067	.5503	.5439
64	.0953	.1015	.7012	.6815	.3779	.3797	.8054	.7969
128	.1868	.1941	.9045	.9002	.6198	.6207	.9634	.9597

by Arias-Castro et al. (2003). They are multiscale, so that the proposed methodology can automatically be adapted to the unknown smoothness of the underlying curve. Note that a faint curve can barely be seen in Figure 6(c).

For each axoid, one considers a statistic that is defined on this axoid. We simply ask: Is this axoid likely to overlap with the underlying curve? If the answer is yes, then this axoid is called *significant*. Two axoids are connected if they can simultaneously cover a geometric curve. (A precise definition of covering was provided in Arias-Castro et al. 2003.) Each axoid can be mapped to a node in a Bernoulli net. Hence the connected significant axoids can be associated with a significance run in the Bernoulli net. The major intuition is that if the image is white noise, then the significant nodes tend to be randomly scattered, and hence the length of the longest significance run tends to be small; however, if there is an embedded curve, then the significant nodes tend to be concentrated around the location of this curve, and hence the length of the longest significance run tends to be large. Based on this intuition, a hypothesis testing scheme can be developed.

Note that the axoids of Huo et al. (2003) and Arias-Castro et al. (2003) may overlap, and hence the derived statistics may be dependent. The assumption that the $X_{i,j}$'s are independent at the beginning of this article is a convenient simplification for obtaining the present results. Extending the current results to the case where the random variables $X_{i,j}$ could be dependent will be an interesting topic for future work.

Our result may have an impact on recent advances in computational vision. Moisan, Desolneux, and Morel (2000) considered how likely it is for some basic geometric objects to be aligned in an image. Only those that are very unlikely to be aligned at random are meaningful to the image content. When the geometric objects can be mapped into a two-dimensional network, the distributional knowledge regarding $L_0(n)$ provides information on how unlikely the observed image is to be generated at random. Hence it provides a way to quantify the threshold of "meaningfulness." Obviously, to apply our results, a substantial amount of formulation and derivation will be required. The idea of using a connectivity pattern in vision research was also explored by, for example, Sha'Ashua and

Table 3. Comparisons of the Simulated $P(L_0 \geq \ell \mid n, 16, 1, p)$ With Approximations by (5) When $n = 32, 64, 128$ and $\ell = 4, 8$

	$p = .15$		$p = .20$		$p = .25$		$p = .30$	
n	S	A	S	A	S	A	S	A
32	.0446	.0474	.2391	.2469	.6116	.6206	.9066	.9018
64	.0948	.1028	.4597	.4692	.8785	.8831	.9952	.9937
128	.1974	.2040	.7200	.7363	.9884	.9889	1.0000	1.0000

Table 4. Comparisons of the Simulated $P(L_0 \geq 8 \mid n, m, 1, p)$ With Approximations by (6)

		$p = .08$		$p = .12$		$p = .16$		$p = .20$	
n	m	S	A	S	A	S	A	S	A
32	32	.0011	.0018	.0223	.0294	.1397	.1502	.4576	.4791
	64	.0033	.0046	.0485	.0683	.2821	.3006	.7121	.7719
	128	.0065	.0102	.0950	.1414	.4840	.5263	.9188	.9563
64	32	.0034	.0042	.0486	.0517	.2867	.2679	.7388	.7549
	64	.0059	.0102	.1057	.1465	.5125	.5184	.9380	.9590
	128	.0128	.0220	.2018	.2889	.7769	.7796	.9970	.9989
128	32	.0070	.0090	.1123	.1321	.5251	.5006	.9435	.9457
	64	.0153	.0212	.2112	.2837	.7789	.7717	.9969	.9987
	128	.0275	.0452	.3795	.5121	.9555	.9523	1.0000	1.0000

Ullman (1988). Moisan et al. (2000) have provided some useful references.

In summary, the results in this article potentially provide a criterion for image feature extraction.

It will also be interesting to derive similar results for a network that is more complicated than a two-dimensional array, for example, a k-dimensional array with certain connectivity conditions, where $k > 2$. One may also be interested in studying a random network in another geometric setting, for example, a connected net of equally spaced points on a sphere.

6.2 Relation to the State of the Art

Over the years there has been considerable research work on the length of the longest success run $L_0(n)$ in n Bernoulli trials, whose extensive applications include signal detection, reliability, quality control, radar astronomy, DNA sequence analysis, startup demonstration testing, and others. Various expressions for the exact distribution of $L_0(n)$ have been given by, among others, Arratia et al. (1990), Balakrishnan and Koutras (2002),

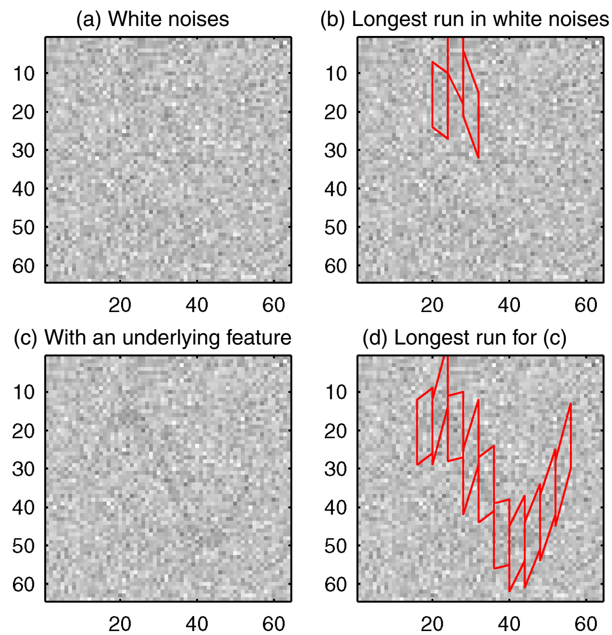

Figure 6. Showcase of Using the Length of the Longest Significance Run to Determine Whether There Is an Embedded Filament. (a) White noise. (b) The corresponding longest significance run. (c) A noisy image with an underlying curve. (d) The corresponding longest significance run.

Burr and Cane (1961), and Gibbons (1971). We found that a summary given by Balakrishnan and Koutras (2002, p. 20) is very helpful. As mentioned earlier, in our formulation, these results are equivalent to the case when $C = 0$ and $m = 1$. In this sense, we generalized the existing results.

Theorem 3 effectively says that the $L_0(n)$ converges to a well-known extreme value distribution. For a quick reference on extreme value distribution, we refer to *http://mathworld.wolfram.com/ExtremeValueDistribution.html*. It is well known [e.g., Fu, Wang, and Lou 2003, eq. (1.5)] that for a one-dimensional Bernoulli sequence, we have

$$P\big(L_0(n) - \lfloor \log_{1/p} n \rfloor < t\big) = \exp\big\{-nqp^{\lfloor \log_{1/p} n \rfloor + t}\big\} + o(1).$$

Historically, it is proven by using the generating function method initiated by Goncharov (1944). Note that when $A_1 = q$, this is a special case of Theorem 3. We found that the literature regarding the limiting distribution of runs in other scenarios has advanced significantly. See Chan and Lai (2003) for a recent inspiring general result.

6.3 Proof Techniques

Our proof of Theorem 3 is based on the Chen–Stein Poisson approximation. There are many ways of using the Poisson approximation. Our approach is the same as that used by Arratia et al. (1990). Notice that there are new developments in this line of methodology; we find that the article of Barbour and Chryssaphinou (2001) provides a good starting point. For us, the method of Arratia et al. (1990) turned out to be sufficient.

7. CONCLUSION

Asymptotic distributions have been derived for the length of the longest significance run in a Bernoulli network. These generalize the known results in longest runs. Efficient numerical algorithms are designed to study the relation between the length of the longest run and the values of parameters in the finite-sample case, as well as the convergence rates to the limit distributions. Our results provide insights in algorithmic designs in applications such as image detection and computational vision.

[Received July 2004. Revised April 2005.]

REFERENCES

Arias-Castro, E., Donoho, D. L., and Huo, X. (2003), "Adaptive Multiscale Detection of Filamentary Structures Embedded in a Background of Uniform Random Points," technical report, Stanford University, available at *http://www-stat.stanford.edu/~donoho/reports.html*; to appear in *The Annals of Statistics*.

Arratia, R., Gordon, L., and Waterman, M. S. (1990), "The Erdös–Rényi Law in Distribution, for Coin Tossing and Sequence Matching," *The Annals of Statistics*, 18, 539–570.

Balakrishnan, N., and Koutras, M. V. (2002), *Runs and Scans With Applications* (2nd ed.), New York: Wiley.

Barbour, A. D., and Chryssaphinou, O. (2001), "Compound Poisson Approximation: A User's Guide," *Annals of Applied Probability*, 11, 964–1002.

Burr, E. J., and Cane, G. (1961), "Longest Run of Consecutive Observations Having a Specified Attribute," *Biometrika*, 48, 461–465.

Chan, H. P., and Lai, T. L. (2003), "Saddlepoint Approximations and Nonlinear Boundary Crossing Probabilities of Markov Random Walks, *Annals of Applied Probability*, 13, 395–429.

Erdös, P., and Rényi, A. (1970), "On a New Law of Large Numbers," *Journal of Analytical Mathematics*, 22, 103–111.

Erdös, P., and Revesz, P. (1975), "On the Length of the Longest Head Run," *Colloqy of the Mathematical Society of Janos Bolyai*, 16, 219–228.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967), "Association of Random Variables, With Applications," *The Annals of Mathematical Statistics*, 38, 1466–1474.

Fu, J. C., Wang, L., and Lou, W. Y. (2003), "On Exact and Large Deviation Approximation for the Distribution of the Longest Run in a Sequence of Two-State Markov-Dependent Trials," *Journal of Applied Probability*, 40, 346–360.

Gibbons, J. D. (1971), *Nonparametric Statistical Inference*, New York: McGraw-Hill.

Goncharov, V. L. (1944), "On the Field of Combinatoric Analysis," *Izvestiya Akademii Nauk SSSR Seriya Matematika*, 8, 3–48 (in Russian); English transl.: *American Mathematical Society Translations*, 19, 1–46.

Huo, X., Chen, J., and Donoho, D. L. (2003), "Multiscale Significance Run: Realizing the 'Most Powerful' Detection in Noisy Images," in *Asilomar Conference on Signals, Systems, and Computers*.

Kulkarni, V. G. (1995), *Modeling and Analysis of Stochastic Systems*, New York: Chapman & Hall.

Moisan, L., Desolneux, A., and Morel, J.-M. (2000), "Meaningful Alignments," *International Journal of Computer Vision*, 40, 7–23.

Muselli, M. (1997), "On Convergence Properties of Pocket Algorithm," *IEEE Transactions on Neural Networks*, 8, 623–629.

——— (2000), "Useful Inequalities for the Longest Run Distribution," *Statistics & Probability Letters*, 46, 239–249.

Papastavridis, S. G., and Koutras, M. V. (1993), "Bounds for Reliability of Consecutive-k-Within-m-Out-of-n Systems," *IEEE Transactions on Reliability*, 42, 156–160.

Petrov, V. V. (1965), "On the Probabilities of Large Deviations for Sums of Independent Random Variables," *Theory of Probability and Its Applications*, 10, 287–298.

Sha'Ashua, A., and Ullman, S. (1988), "Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network," in *Second International Conference on Computer Vision*, pp. 321–327.

Wallenstein, S., Naus, J., and Glaz, J. (1994), "Power of the Scan Statistic in Detecting a Changed Segment in a Bernoulli Sequence," *Biometrika*, 81, 596–601.