The Annals of Probability
2003, Vol. 31, No. 4, 1713–1753

RIGOROUS RESULTS FOR THE NK MODEL

BY RICHARD DURRETT[1] AND VLADA LIMIC[2]

Cornell University

Motivated by the problem of the evolution of DNA sequences, Kauffman and Levin introduced a model in which fitnesses were assigned to strings of 0's and 1's of length N based on the values observed in a sliding window of length $K + 1$. When $K \geq 1$, the landscape is quite complicated with many local maxima. Its properties have been extensively investigated by simulation but until our work and the independent investigations of Evans and Steinsaltz little was known rigorously about its properties except in the case $K = N - 1$. Here, we prove results about the number of local maxima, their heights and the height of the global maximum. Our main tool is the theory of (substochastic) Harris chains.

1. Introduction. In Kauffman and Levin's (1987) NK model, N refers to the number of parts of the system—genes in a genome, amino acids in a protein, nucleotides in a DNA sequence—and each part makes a contribution to the overall fitness that depends on that part and on K other parts among the N. To have the simplest possible setting, we will suppose that each part has two possible states and represent the state of the system by $\eta \in \{0, 1\}^{\{0,1,\ldots,N-1\}}$. The fitness of η is

$$\Phi(\eta) = \sum_{i=0}^{N-1} \phi_i(\eta_i, \ldots, \eta_{i+K}), \tag{1.1}$$

where the arithmetic in the subscripts is done modulo N and the $\phi_i(\eta_i, \ldots, \eta_{i+K})$ are i.i.d. with a distribution function $F(x) = \int_{-\infty}^{x} f(y)\,dy$. To simplify the proofs and to have only one set of hypotheses, we will suppose throughout the paper that the density function f is continuous on the interior of its support and that

$$\int e^{\theta x} f(x)\,dx < \infty \qquad \text{for } \theta \in (-\delta, \delta) \text{ for some } \delta > 0. \tag{F}$$

The main motivation for assuming the existence of a density is to make use of thetheory of Harris chains. Weaker assumptions would suffice for many results, but our stronger assumptions cover all the examples that have been studied previously.

Kauffman and Levin (1987), and much of the work that followed, focused on the special case in which F is uniformly distributed on $(0, 1)$. This is the most important special case. However, we will consider the model with general F, since

Received August 2001; revised July 2002.

[1]Supported in part by NSF Grant DMS-98-77066.

[2]Supported in part by an NSF postdoctoral fellowship and NSF Grant DMS-01-04232.

AMS 2000 subject classifications. 60G50, 60F05.

Key words and phrases. NK model, fitness, local maxima, limit theorems, R-recurrence.

it is interesting as well. Weinberger (1991) performed a physicist's analysis of the case in which F has the standard normal distribution. Evans and Steinsaltz (2001) calculated various quantities of interest for the NK model, where F is the (positive) exponential distribution $F(x) = 1 - e^{-x}$, $x \geq 0$, or the Gamma(2, 1) distribution with density $f(x) = xe^{-x}$. In this paper, we prove analogous detailed results when F is the negative exponential distribution $F(x) = e^x$ for $x \leq 0$.

As Kauffman and Levin (1987) observed, the case $K = 0$ is trivial. The parts do not interact, so there is only one maximum $(\eta_0^*, \dots, \eta_{N-1}^*)$, which is obtained by choosing η_i^* to maximize $\eta_i \mapsto \phi_i(\eta_i)$ for each i. The other extreme $K = N - 1$ is also simple. Each ϕ_i is a function of all N coordinates, so the fitness of each η is a sum of N independent uniforms and the values of $\Phi(\eta)$ are independent for different η. The probability that a vertex is a local maximum is just the probability it is larger that its N neighbors, $1/(N + 1)$, and thus the expected number of local maxima is $EM_N = 2^N/(N + 1)$. Other aspects of the fully interconnected case $K = N - 1$ lead to some interesting questions. Kauffman and Levin (1987) argued heuristically that if one starts at a randomly chosen point and moves to a fitter neighbor chosen at random, then the adaptive walk takes an average of $\log_2 N$ steps to reach a local maximum. Weinberger (1988), Macken and Perelson (1989), Macken, Hagan and Perelson (1991) and Flyvberg and Lautrup (1992) carried out more detailed analysis of such walks. See Chapter 2 of Kauffman (1993) for other quantities related to the $N = K - 1$ landscape that have been analyzed.

We now describe the contents of this paper in some detail. The precise statements and proofs of all theorems stated in this section can be found in Sections 2–7. Let $G_N = \{(0, 0, \dots, 0)$ is a local maximum for $\Phi\}$ and note that $EM_N = 2^N P(G_N)$. Our first result implies that, for a fixed K, the quantity $(\log EM_N)/N$ converges to a limit.

THEOREM 2.1. *For each fixed $K \geq 1$, there is a constant λ_K so that*

$$\lim_{N\to\infty} \frac{1}{N} \log P(G_N) = \log \lambda_K.$$

Note that, for all K and F, we have $\lambda_K \geq 1/2$ since $EM_N \geq 1$, and also $\lambda_K < 1$, due to the sentence following Theorem 5.1.

REMARK. In (1.10), we will show that $P(G_N) \sim C\lambda_K^N$. Evans and Steinsaltz (2002), see their Theorem 7, have shown this result for distributions that are bounded below.

Theorem 2.1 is a simple consequence of subadditivity. For $1 \leq i \leq N$, let 0_i be the vector that has a 1 at coordinate i and 0 at all other coordinates, and let E_i be the event that changing the bit at i from 0 to 1 does not increase the value:

$$E_i = \{\phi_{i-K}(0) + \dots + \phi_i(0) > \phi_{i-K}(0_i) + \dots + \phi_i(0_i)\}. \tag{1.2}$$

Here, and in what follows, it is convenient to regard ϕ_j as a function of the entire sequence by setting $\phi_j(\eta) = \phi_j(\eta_j, \dots, \eta_{j+K})$. Let $G'_{N-1} = \bigcap_{i=K}^{N-1} E_i$ be the part of G_N that involves tests that do not "wrap around." Clearly, $P(G_N) \le P(G'_{N-1})$. It is not hard to show that $P(G'_{N-1}) \le P(G'_{M-1})P(G'_{N-M-1})$ and hence

$$\frac{1}{N} \log P(G'_{N-1}) \to \inf_M \frac{1}{M} \log P(G'_{M-1}).$$

The proof of Theorem 2.1 can then be completed (see Section 2 for details) by showing $P(G_N) \ge \varepsilon_K P(G'_{N-1})$.

As is usually the case in applications of subadditivity, the proof of Theorem 2.1 gives no insight into the value of the constant, except for the crude upper bounds that come from the definition of the limiting constant. In Section 3, we study the case of the negative exponential distribution $F(x) = e^x$, $x \le 0$, in detail. In particular, we compute λ_1 exactly. If we change variables $y_i = -x_i$, a formula of Weinberger (1991) implies that (if $K = 1$)

$$(1.3) \quad P(G_N) = \int_0^\infty dy_0 \cdots \int_0^\infty dy_{N-1} \exp\left(-\sum_{i=0}^{N-1} 3y_i\right) \prod_{i=0}^{N-1} (1 + y_i + y_{i-1}).$$

After integrating out y_0 to break the ring, one can write recursive equations for related multiple integrals to conclude that

$$\frac{1}{N} \log P(G_N) \to \log\left(\frac{5 + \sqrt{29}}{18}\right),$$

so $\lambda_1 \approx 0.5769536$. Replacing 3 by $3 + \theta$ in (1.2) gives a formula for the Laplace transform of the sum of the coordinates on G_N:

$$\begin{aligned}\hat{Z}_N(\theta) &= E\left(\exp\left(-\theta \sum_{i=0}^{N-1} y_i\right); G_N\right) \\ &= \int_0^\infty dy_0 \cdots \int_0^\infty dy_{N-1} \exp\left(-\sum_{i=0}^{N-1} (3+\theta) y_i\right) \prod_{i=0}^{N-1} (1 + y_i + y_{i-1}).\end{aligned}$$

We use the notation $\hat{Z}_N(\theta)$ to indicate that this is the analogue of the partition function from statistical mechanics. Using the recursions to compute the Laplace transform and then differentiating, we find that

$$\lim_{N\to\infty} E\left(\frac{1}{N} \sum_{i=1}^{N} x_i \middle| G_N\right) = -\frac{126\sqrt{29} + 774}{270\sqrt{29} + 1566} = -0.480971328.$$

This gives the expected height of a local maximum at 0 conditioned on G_N.

Differentiating again, we can find the asymptotic distribution of the second moment and use the formula for the Laplace transform to obtain the following result.

THEOREM 3.1. *If F is the negative exponential distribution and $K = 1$, then the distribution of $(\phi(0) - \mu_H N)/\sqrt{N}$ conditional on G_N converges to a normal with mean 0 and variance σ_H^2. Here, μ_H is the mean given above and $\sigma_H^2 \approx 0.901465824$ is a constant that has an exact formula similar to the one for μ_K.*

An earlier version of this paper asserted that the limit law in Theorem 3.1 applied to the height of a randomly chosen local maxima since that quantity had the same distribution as $\phi(0)$ conditional on G_N. Unfortunately, this is a statement that is "obviously true" but turns out to be false. Fix N, K and recall that M_N denotes the number of local maxima in the landscape. Let I_N be the coordinates of a randomly chosen local maximum. Using the symmetry of the landscape, we have $P(I_N = (0, 0, \dots, 0)|M_N = m) = 2^{-N}$ and $P(I_N = (0, 0, \dots, 0)) = 2^{-N}$, so

$$\begin{aligned}
&P\bigl(M_N = m|I_N = (0, 0, \dots, 0)\bigr)\\
&\quad = \frac{P(I_N = (0, 0, \dots, 0)|M_N = m)P(M_N = m)}{P(I_N = (0, 0, \dots, 0))}\\
&\quad = \frac{2^{-N}P(M_N = m)}{2^{-N}} = P(M_N = m).
\end{aligned}$$

In words, if we pick a local maximum at random and its coordinates turn out to be all 0's, then the distribution of M_N is not changed. In contrast, since $P(G_N|M_N = m) = m/2^N$, we have

$$\begin{aligned}
P(M_N = m|G_N) &= \frac{P(G_N|M_N = m)P(M_N = m)}{P(G_N)}\\
&= \frac{P(M_N = m)m/2^N}{P(G_N)},
\end{aligned}$$

so conditioning the landscape on G_N causes it to have more local maxima. The shift caused by the weight factor $m/2^N$ is far from innocent. Theorem 7.1 will show that $(\log M_N - \mu_M N)/\sqrt{N}$ has a limiting normal distribution. The last computation is not definitive, but once one doubts the obvious result is true, it is easy to see it fails even in the simplest possible example. Suppose $N = 2$ and $K = 1$. Since $K = N - 1$, the four heights $\Phi(i, j)$ are independent. The probability density

$$P\bigl(\Phi(0, 0) = h, G_2\bigr) = P\bigl(\Phi(0, 0) = h > \max\{\Phi(1, 0), \Phi(0, 1)\}\bigr) = f(h)F(h)^2.$$

On the other hand,

$$\begin{aligned}
&P\bigl(\Phi(0, 0) = h, I_2 = (0, 0)\bigr)\\
&\quad = \tfrac{1}{2}P\bigl(\Phi(0, 0) = h, \min\{h, \Phi(1, 1)\} > \max\{\Phi(1, 0), \Phi(0, 1)\}\bigr)\\
&\qquad + P\bigl(\Phi(0, 0) = h > \max\{\Phi(1, 0), \Phi(0, 1)\} > \Phi(1, 1)\bigr).
\end{aligned}$$

Break things down according to the value of $\Phi(1,1)=x$ to get

$$= \tfrac{1}{2} f(h)F(h)^2\bigl(1-F(h)\bigr) + \tfrac{1}{2} f(h)\int_{-\infty}^{h} dx\, f(x)F(x)^2 + f(h)\int_{-\infty}^{h} dx\, f(x)\bigl(F(h)^2 - F(x)^2\bigr).$$

The ratio of the two densities just computed is not constant, so

$$P\bigl(\Phi(0,0)=h|G_2\bigr) \neq P\bigl(\Phi(0,0)=h|I_2=(0,0)\bigr).$$

Our next result considers the height of the global maximum, H_N^*.

THEOREM 3.2. *Suppose F has a negative exponential distribution. Let $b = -0.231961$. If $a > b$, then $\sup_{K<N} P(H_N^* > aN) \to 0$ as $N \to \infty$.*

This is proved by using standard large-deviations estimates for sums of random variables. It is sharp in the case $K_N = N-1$ since, then, the values at the 2^N points are independent. It is certainly not sharp when $K=0$ and is presumably an overestimate for other fixed values of K.

Evans and Steinsaltz (2001) have studied λ_K, H_N and H_N^* for the positive exponential distribution $F(x) = 1 - e^{-x}$, $x \geq 0$. Their results show that, when $K=1$, $\lambda_1 = 0.562682$, $H_N/N \to 1.61651$ and $H_N^*/N \to 1.78509$. The first two results could also be derived using the methods of Section 3. Theorem 3.3 gives an upper bound of 2.678347 on the limit of H_N^*/N, so we suspect that our bound for the negative exponential from Theorem 3.2 is not very good, either, when $K=1$. Theorems 6.2 and 7.2 improve the results of Evans and Steinsaltz by giving central limit theorems for H_N and H_N^* for the positive exponential.

We are not able to get exact results for λ_1 in the uniform case, but we are able to get reasonably good bounds. To explain this, we will introduce a connection with Markov chains that is valid for a general F, N and $K<N$ and will be the key to many of our theoretical results. Recall that E_i is the event that changing the bit at i from 0 to 1 does not increase the overall fitness. Let $X_j = \phi_j(0)$ and let $\mathcal{F}_k$ be the σ-field generated by $\phi_j(0)$ and $\phi_j(0_i)$ with $i, j \leq k$. The definitions (1.1) and (1.2) imply that $E_j \in \mathcal{F}_k$ if $K \leq j \leq k$, and if $k \geq K-1$, then, on $G_k' = \bigcap_{i=K}^{k} E_i$,

$$P(E_{k+1}, X_{k+1}=y|\mathcal{F}_k) = F_{K+1}(X_{k-K+1} + \cdots + X_k + y) f(y), \tag{1.4}$$

where F_K is the distribution of the sum of K independent random variables with distribution F. The last equation shows that X_j is a K-step Markov process; that is, $(X_{j-K+1}, \dots, X_j)$, $j \geq K-1$, is a Markov chain.

When $K=1$ and F is uniform, (1.4) states

$$p(x,y) = P(E_{k+1}, X_{k+1}=y|X_k=x, \mathcal{F}_k) = F_2(x+y), \qquad x, y \in [0,1].$$

Since p is a symmetric and square-integrable function, a theorem on page 243 of Riesz and Nagy (1990) implies that we can write

$$p(x, y) = \sum_{i=1}^{\infty} \beta_i h_i(x) h_i(y), \tag{1.5}$$

where β_i is a decreasing sequence of eigenvalues and the $h_i(x)$ are the corresponding eigenfunctions which form an orthonormal sequence. In Section 4, we establish that $\lambda_1 = \beta_1$ and obtain bounds on λ_1.

To get a lower bound on λ_1, we can use the variational characterization of the largest eigenvalue

$$\beta_1 = \max \frac{\iint g(x) p(x, y) g(y)\, dx\, dy}{\int g(x)^2\, dx}.$$

A little calculus shows that if we choose $g(x) = 1 + ax$ and then optimize the value of a, we have

$$\lambda_1 \geq 0.571455. \tag{1.6}$$

To get a bound in the other direction, let

$$q_N(x) = P(G'_{N-1} | X_0 = x).$$

As one might expect [see inequality (2.10)], $q_N(x)$ is the largest for $x = 1$. Another application of subadditivity implies

$$(1/N) \log q_N(1) \to \inf_{M \geq 1} (1/M) \log q_M(1) = \log \lambda_1.$$

Using Mathematica to compute $q_5(1) = 0.0839578$ then gives

$$\lambda_1 \leq q_5(1)^{1/5} = 0.60273. \tag{1.7}$$

As the referee pointed out, one might be able to estimate λ_1 by approximating the continuous eigenvalue problem by a numerical eigenvalue problem for a large matrix. We leave this project to an interested reader.

In Section 5, we study the behavior of λ_K for large K. Recall that throughout the paper we are assuming F is a distribution satisfying (F).

THEOREM 5.1. *For large* K,

$$\lambda_K \geq 1 - \frac{9 \log(K+1)}{K+1}.$$

For a corresponding upper bound, one can note that

$$P(G_N) \leq P\left(\bigcap_{i=1}^{[N/(K+1)]} E_{i(K+1)} \right) = (1/2)^{[N/(K+1)]}$$

since the events are independent, so $\lambda_K \leq (1/2)^{1/K+1} \approx 1 - (\ln 2)/(K+1)$. We believe that the bound in Theorem 5.1 is sharp, that is, we have the following.

CONJECTURE. *There is a* $c > 0$ *so that* $\lambda_K \le 1 - c\log(K+1)/(K+1)$.

In support of this conjecture, note that if the actual number of local maxima (not just its expected value) is of order $2^N(1 - c\log(K+1)/K+1)^N$, then a large-deviations calculation would show that if the mean of ϕ_i is 0 and the variance is 1, then the average height of the local maxima would be less than or equal to $C\sqrt{\log(K+1)/(K+1)}$, in agreement with the heuristic calculations of Weinberger (1991); see his page 6401. To see why $C\sqrt{\log(K+1)/(K+1)}$ is a reasonable guess for the limit of EH_N/N, suppose that we divide the coordinates into blocks of size $K+1$ and the fitness contribution of each site in each block depends on the $K+1$ coordinates in the block. In this case, the fitness contribution of each block will have approximately a normal distribution with mean 0 and variance $K+1$. Each block behaves like the fully interconnected case so, as discussed above, a local maximum will look like the maximum of $K+1$ independent normals with mean 0 and variance $K+1$, which will have mean approximately equal to $\sqrt{2(K+1)\log(K+1)}$. [See, e.g., Exercise 2.3 on page 85 of Durrett (1995).]

In Sections 6 and 7, we study the general model where $K \ge 1$ and F satisfying (F) are fixed. We consider the vector Markov chain with transition probability

$$q\big((x_0,\dots,x_{K-1}),(x_1,\dots,x_K)\big) = F_{K+1}(x_0+\cdots+x_K)f(x_K)$$

and is 0 for other choices of the second argument. In Section 6, we show that this chain is R-recurrent in the sense of Tweedie (1974). Let

$$Q(y,A) = \int_A q(y,z)\,dz.$$

Results in Section 3 of Tweedie (1974) now imply the existence of a constant R, a measure μ and a function h unique up to constant multiples so that

$$\int \mu(dy)RQ(y,A) = \mu(A) \quad \text{and} \quad \int RQ(y,dz)h(z) = h(y). \tag{1.8}$$

Theorem 6 on page 860 of Tweedie (1974) then implies that

$$R^nQ^n(x,A) \to \frac{\mu(A)h(x)}{\int \mu(dy)h(y)}. \tag{1.9}$$

From this, we can conclude easily that

$$P(G_N) \sim C/R^N, \tag{1.10}$$

sharpening the conclusion of Theorem 2.1.

To investigate the properties of coordinates of local maxima, it is useful to let $\pi(y) = cj(y)$, where $d\mu(y) = j(y)\,dy$ and where the constant c is chosen so that $\int dy\,\pi(y)h(y) = 1$. Introduce the transformed chain

$$\bar q(x,y) = \frac{R}{h(x)}q(x,y)h(y). \tag{1.11}$$

Since $h(y)$ is a right eigenvector, the kernel $\bar{q}$ satisfies has $\int \bar{q}(x, y)\, dy = 1$. Since $\pi(x)$ is a left eigenvector, $\bar{\pi}(x) = \pi(x)h(x)$ is a stationary distribution

$$\int dx\, \pi(x)h(x)\bar{q}(x, y) = \pi(y)h(y).$$

Let P_N be the distribution of $(x_0, \dots, x_{N-1})$ conditioned on G_N and let Q_N be the distribution of $(x_0, \dots, x_{N-1})$ under the Markov chain with transition probability $\bar{q}$ and initial distribution $\bar{\pi}$. Results from the theory of Markov chains give limit theorems under Q_N. Considering the Radon–Nikodym derivative dP_N/dQ_N then allows us to transfer the well-known results from Q_N to P_N.

THEOREM 6.1. *Let* $\mu_H = \int dy\, \pi(y)h(y)\, y$. *If* $\varepsilon > 0$, *then*

$$P\big(|\phi(0)/N - \mu_H| > \varepsilon | G_N\big) \to 0.$$

THEOREM 6.2. *There is a constant* σ_H^2 *so that the distribution of* $(\phi(0) - \mu_H N)/\sqrt{N}$ *conditional on* G_N *converges to a normal with mean* 0 *and variance* σ_H^2.

As in the case of Theorem 3.1, this shows that most of the local maxima have about the same height. Figure 1 shows the result of 1000 simulations of a system with $N = 96$ and $K = 1$. The distribution has a roughly normal shape but is somewhat asymmetric. The reader who complains that 96 is not very large should note that there are $2^{96} > 10^{28}$ points in the space.

In Section 7, we prove results about the number of local maxima and the height of the global maximum. The key to this is the observation that there are "cut points" where all local maxima must have specified bits and this breaks the overall maximization problem into a large number of independent maximization subproblems. To explain this notion, consider the special case $K = 1$. If

$$\phi_{i-1}(u, 1) + \phi_i(1, v) > \phi_{i-1}(u, 0) + \phi_i(0, v)$$

for the four choices of $u, v \in \{0, 1\}$, then the ith coordinate of any local maximum must be 1, and we call i a cut point. If $j > i$ is another cut point, then the optimization inside the segment (i, j) can be done independently of the other variables.

Combining the idea of cut points with results about Harris chains, it is easy to show the following result.

THEOREM 7.1. *Let* M_N *be the number of local maxima. There are constants* μ_M *and* σ_M^2 *such that* $(\log M_N - \mu_M N)/\sqrt{N}$ *converges in distribution to a normal with mean* 0 *and variance* σ_M^2.

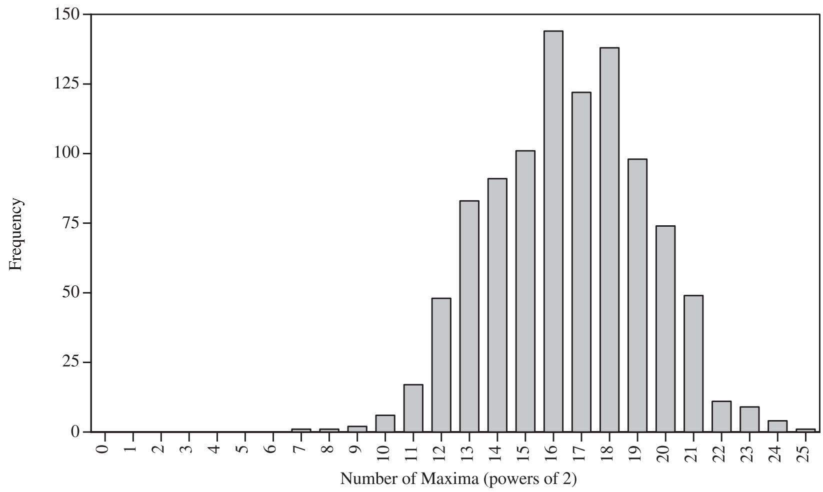

FIG. 1. *Number of maxima (powers of 2).*

THEOREM 7.2. *Let H_N^* be the height of the global maximum. There are constants μ_{H^*} and $\sigma_{H^*}^2$ such that $(H_N^* - \mu_{H^*}N)/\sqrt{N}$ converges in distribution to a normal with mean* 0 *and variance $\sigma_{H^*}^2$.*

Figures 2 and 3 show the results of 1000 simulations of the quantities studied in Theorems 7.1 and 7.2 for $N = 96$ and $K = 1$. Note the wide range for the number of maxima from $2^7 = 128$ to over 2^{25}, which is approximately 32 million. The shape of both distributions is decidedly abnormal.

Up to this point, we have been concerned with the height of $\Phi(0, 0, \dots, 0)$ given $(0, 0, \dots, 0)$ is a local maximum. Theorem 7.3 shows that the ensemble of values in one realization is approximately normal for large N. That is, if ν_N is the measure

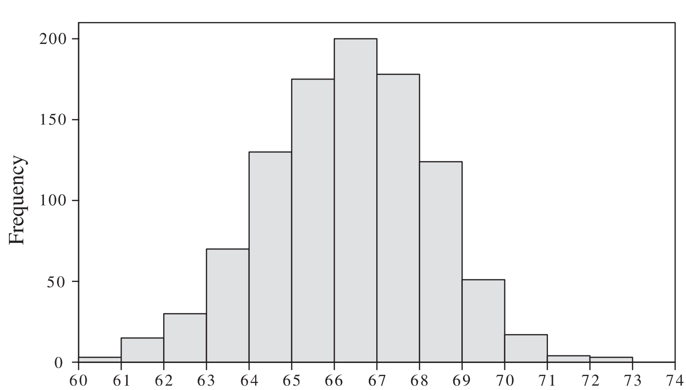

FIG. 2. *Randomly chosen local maximum.*

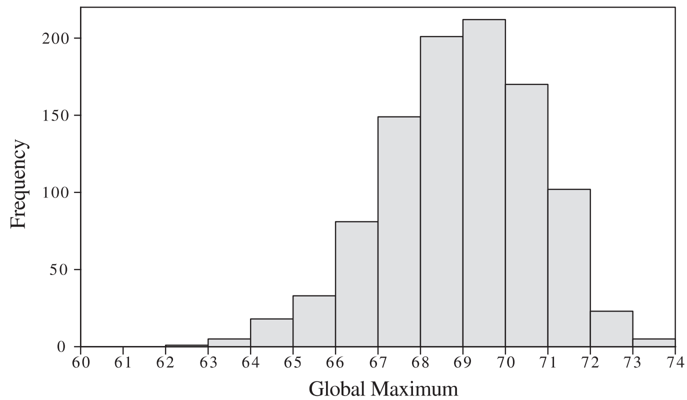

FIG. 3. *Global maximum.*

that assigns mass $1/M_N$ to the height of each local maximum and N is large, then ν_N has approximately a normal distribution. Figure 4 gives the heights of local maxima in one simulation of the system with $N = 96$ and $K = 1$. The distribution has almost exactly the shape of the normal distribution. Since the mode of the center of the distribution is $O(N)$ while the standard deviation is $O(\sqrt{N})$, this leads to the interesting qualitative conclusion that most of the local maxima have about the same height.

The results in Section 7 also give some insight into what Kauffman calls the Massif Central phenomenon: local optima with high values are close to the global

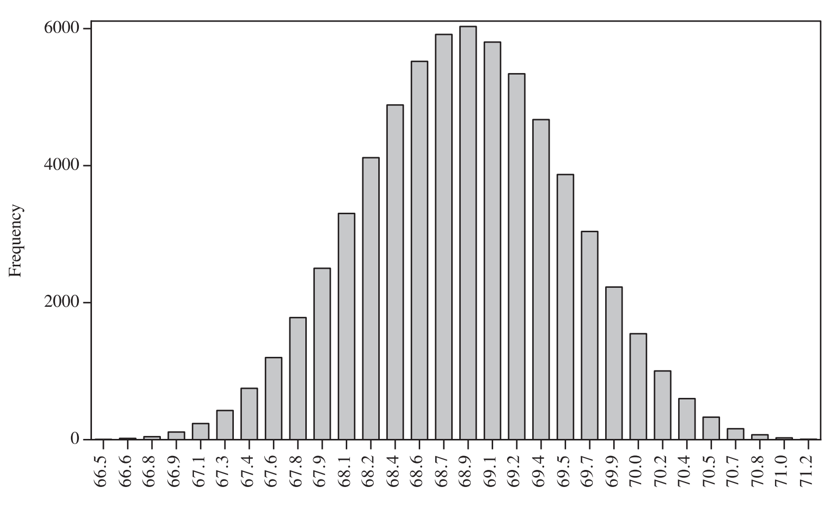

FIG. 4. *Height.*

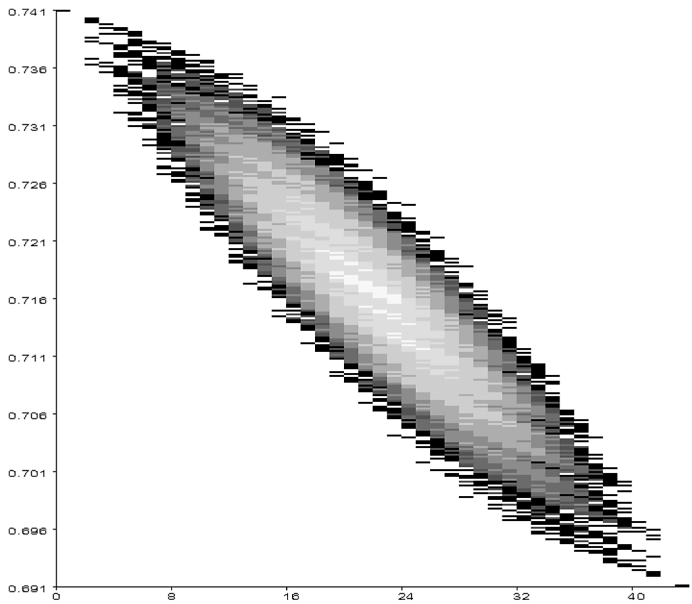

FIG. 5.

maximum in the usual metric on the hypercube defined by

$$d(\eta^1, \eta^2) = \sum_{i=0}^{N-1} |\eta_i^1 - \eta_i^2|$$

for $\eta^1, \eta^2 \in \{0, 1\}^{\{0,1,\dots,N-1\}}$. This is illustrated in the simulation of $N = 256$ and $K = 1$ in Figure 5. Here, the 69,578,335,677,472 local maxima are classified according to their height and distance from the global maximum. Each band represents an increase in density by a factor of 16. Two randomly chosen points have a distance that is binomial with mean 128 and standard deviation 8. However, in the simulation no local maximum is at distance more than 120 from the global maximum and the typical local maximum has a distance between 40 and 75.

To understand this phenomenon, intuitively we note that the cut points break the maximization problem into independent pieces. A solution that does not make the best local choice in a positive fraction of the intervals will be smaller than the global maximum by a constant times N. Conversely, those local maxima whose heights are within εN of the global maximum must be close to it. It would be interesting to prove results about the limiting shape of the picture in Figure 5 and the limiting behavior of $(1/N \log M_N(c))$, where $M_N(c)$ is the number of local maxima at distance $[cN]$ from the global maximum. The simulation supports the notion that such a limit exists. However, at this point, we do not know how to attack these two large-deviations problems.

2. General results. Let $G_N = \{(0,0,\dots,0) \in \mathbf{R}^N$ is a local maximum for $\Phi\}$.

THEOREM 2.1. *There is a constant λ_K so that*

$$\lim_{N\to\infty} \frac{1}{N} \log P(G_N) = \log \lambda_K.$$

PROOF. For $1 \le i \le N$, let 0_i be the vector that has a 1 at the ith coordinate and 0 at all other coordinates, and let

$$E_i = \{\phi_{i-K}(0) + \cdots + \phi_i(0) > \phi_{i-K}(0_i) + \cdots + \phi_i(0_i)\}$$

so that $G_N = \bigcap_{k=1}^N E_k$. Let $V_i = \phi_i(0)$ and $V_{ji} = \phi_j(0_i)$ when $i - K \le j \le i$. Since the random variables V_i and V_{ji} are independent, after conditioning on the values of $V_0, \dots, V_{N-1}$ we arrive at a special case of formula (2.4) of Weinberger (1991):

$$P(G_N) = \int dF(v_0) \cdots \int dF(v_{N-1}) \prod_{i=0}^{N-1} F_{K+1}\left(\sum_{j=i-K}^{i} v_j\right), \tag{2.1}$$

where F_{K+1} is the distribution function of the sum of $K+1$ random variables with distribution F, and summation is modulo N.

To prove Theorem 2.1, we will first consider $G'_{N-1} = \bigcap_{i=K}^{N-1} E_k$, which leaves out the terms that "wrap around," and show that

$$\lim_{N\to\infty} \frac{1}{N} \log P(G'_{N-1}) = \log \lambda_K. \tag{2.2}$$

Clearly, $P(G_N) \le P(G'_{N-1})$. To bridge the gap, we soon show that $P(G_N) \ge \varepsilon_K P(G'_{N-1})$. First, observe that $P(G'_{N-1})$ is a submultiplicative sequence, that is,

$$\begin{aligned} P(G'_{N-1}) \le \int dF(v_0) \cdots \int dF(v_{N-1}) \prod_{i=K}^{M-1} F_{K+1}\left(\sum_{j=i-K}^{i} v_j\right) \\ \times \prod_{i=M+K}^{N-1} F_{K+1}\left(\sum_{j=i-K}^{i} v_j\right) \\ = P(G'_{M-1}) P(G'_{N-M-1}). \end{aligned}$$

A standard subadditivity argument now shows that

$$\frac{1}{N} \log P(G'_{N-1}) \to \inf_{M\ge 1} \frac{1}{M} \log P(G'_{M-1}),$$

and we have established (2.2).

To complete the proof of Theorem 2.1, we note that $W_i^N = F_{K+1}(\sum_{j=i-K}^{i} V_j)$ are increasing functions of independent random variables so Harris's inequality, see, for example, Kesten (1981), implies that they have positive correlations. The superscript N is here to remind us that since we use modulo arithmetic, the values of the W_i^N depend on i and N. Harris's inequality implies that

$$P(G_N) \geq E\left(\prod_{i=0}^{K-1} W_i^N\right) P(G'_{N-1}). \tag{2.3}$$

The first term on the right-hand side is a positive quantity whose value does not depend on N, so the proof is complete. □

To further investigate properties of $P(G'_{N-1})$, we will introduce a K-step Markov process. Let $G'_j = \bigcap_{i=K}^{j} E_i$ (with $G'_j = \Omega$ if $j < K$) and let

$$X_j = \begin{cases} \phi_j(0), & \text{on } G'_j, \\ \Delta, & \text{on } (G'_j)^c, \end{cases}$$

where Δ is a cemetery state that indicates 0 is not a local maximum. Let $\mathcal{F}_k$ be the σ-field generated by $\phi_j(0)$ and $\phi_j(0_i)$ with $i, j \leq k$. The definitions (1.1) and (1.2) imply that $E_j \in \mathcal{F}_k$ if $K \leq j \leq k$, and if $k \geq K - 1$, then, on G'_k,

$$P(E_{k+1}, X_{k+1} = y|\mathcal{F}_k) = F_{K+1}(X_{k-K+1} + \cdots + X_k + y) f(y). \tag{2.4}$$

The last equation shows that X_j is a K-step Markov process; that is, $(X_{j-K+1}, \ldots, X_j)$, $j \geq K - 1$, is a Markov chain.

The first K values, $X_0, \ldots, X_{K-1}$, are the initial condition for the Markov chain. If we let

$$p(y|x_{K-1}, \ldots, x_0) = f(y) F_{K+1}(x_0 + \cdots + x_{K-1} + y),$$

then iterate and use (2.4), we have

$$P(G'_{N-1}) = \int dF(x_0) \cdots \int dF(x_{K-1})\, q_{N-K}(x_0, \ldots, x_{K-1}), \tag{2.5}$$

where

$$q_{N-K}(x_0, \ldots, x_{K-1}) = \int dx_K \cdots \int dx_{N-1} \prod_{j=K}^{N-1} p(x_j|x_{j-1}, \ldots, x_{j-K}).$$

When $K = 1$, X_j is a Markov chain with transition probabilities

$$P(X_1 = dy|X_0 = x) = p(x, y)\, dy = F_2(x + y) f(y)\, dy, \qquad y \neq \Delta,$$

$$P(X_1 = \Delta|X_0 = x) = 1 - \int F_2(x + y) f(y)\, dy,$$

$$P(X_1 = \Delta|X_0 = \Delta) = 1.$$

If we define

$$\bar{p}(x, y) = f(x)^{1/2} p(x, y) f(y)^{-1/2},$$

then $\bar{p}(x, y) = \bar{p}(y, x)$ for $x, y \neq \Delta$ and

$$\bar{p}^n(x, y) = f(x)^{1/2} p^n(x, y) f(y)^{-1/2}. \tag{2.6}$$

When $K > 1$, the Markov chain $(X_{j-K+1}, \dots, X_j)$ has transition probabilities

$$q\big((x_0, \dots, x_{K-1}), (x_1, \dots, x_K)\big) = F_{K+1}(x_0 + \dots + x_K) f(x_K),$$

$$q\big((x_0, \dots, x_{K-1}), \Delta\big) = 1 - \int_{-\infty}^{\infty} f(x) F_{K+1}(x_0 + \dots + x_{K-1} + x)\, dx.$$

This chain has the symmetry property

$$\begin{aligned} &f(x_0)^{1/2} q\big((x_0, \dots, x_{K-1}), (x_1, \dots, x_K)\big) f(x_K)^{-1/2} \\ &\quad = f(x_K)^{1/2} q\big((x_K, \dots, x_1), (x_{K-1}, \dots, x_0)\big) f(x_0)^{-1/2}. \end{aligned} \tag{2.7}$$

This is similar to, but not quite, the reversibility and seems much weaker than the self-adjointness that holds when $K = 1$.

Still the K-step chain has nice monotonicity properties. If $x_i \geq x_i'$ for $0 \leq i \leq K - 1$, then

$$\begin{aligned} &q\big((x_j, \dots, x_{K-1}, z_K, \dots, z_{j+K-1}), (x_{j+1}, \dots, x_{K-1}, z_K, \dots, z_{j+K})\big) \\ &\quad \geq q\big((x_j', \dots, x_{K-1}', z_K, \dots, z_{j+K-1}), (x_{j+1}', \dots, x_{K-1}', z_K, \dots, z_{j+K})\big). \end{aligned} \tag{2.8}$$

By iterating (2.8), we get that if $x \geq x'$ coordinatewise and $n \geq K$, then, for all $x, y \in \mathbf{R}^K$,

$$q^n(x, y) \geq q^n(x', y). \tag{2.9}$$

After integrating, we have

$$x \to q_n(x) = \int_{\mathbf{R}^K} q^n(x, y)\, dy \qquad \text{is increasing.} \tag{2.10}$$

This holds for $n \geq K$ by (2.9). Using (2.8), we see that it is also valid for $1 \leq n \leq K - 1$. If we let $q_n^* = \sup_x q_n(x)$, then it is easy to see that $q_n^* \leq q_m^* \cdot q_{n-m}^*$ and, hence, that

$$\frac{1}{n} \log q_n^* \to \inf_{m \geq 1} \frac{1}{m} \log q_m^* = \log \lambda_K. \tag{2.11}$$

To explain the last equality, observe

$$P(G_{N-1}') = \int F(dx_0) \cdots \int F(dx_{K-1}) \int_{\mathbf{R}^K} q^{N-K}(x, y)\, dy \leq q_{N-K}^* \tag{2.12}$$

and

$$\begin{aligned} q_N(x) &= \int_{\mathbf{R}^K} \int_{\mathbf{R}^K} q^K(x, z) q^{N-K}(z, y)\, dz\, dy \\ &\leq \int F(dz_0) \cdots \int F(dz_{K-1}) \int q^{N-K}(z, y)\, dy = P(G_{N-k}'), \end{aligned}$$

so $q_N^* \leq P(G_{N-K}')$.

3. Results for the negative exponential. Consider now the case in which $F(x) = e^x$ for $x \le 0$; that is, $-X$ has exponential (rate 1) distribution. We begin with the case $K = 1$. If X_1 and X_2 are independent and have distribution F, then $P(X_1 + X_2 \le -t)$ is the probability that there have been 0 or 1 arrivals in a rate 1 Poisson process at time t, that is, $e^{-t}(1+t)$. Changing variables $t = -v$, we have $F_2(v) = e^v(1-v)$ for $v \le 0$. Using (2.1) and changing variables $x_i = -v_i$, we have

$$P(G_N) = \int_0^\infty dx_0 \cdots \int_0^\infty dx_{N-1} \exp\left(-\sum_{i=0}^{N-1} 3x_i\right) \prod_{i=0}^{N-1} (1 + x_i + x_{i-1}). \tag{3.1}$$

To get more information, with only a little extra work, we will analyze the Laplace transform

$$\begin{aligned}\hat{Z}_N(\theta) &= E\left(\exp\left(-\theta \sum_{i=0}^{N-1} x_i\right); G_N\right) \\ &= \int_0^\infty dx_0 \cdots \int_0^\infty dx_{N-1} \exp\left(-\sum_{i=0}^{N-1} (3+\theta)x_i\right) \prod_{i=0}^{N-1} (1 + x_i + x_{i-1}).\end{aligned}$$

We use the notation $\hat{Z}_N(\theta)$ to indicate that this is the analogue of the partition function from statistical mechanics.

To analyze this expression, it is useful to let $\Pi_a^b = \Pi_a^b(x_0, x_1, \ldots, x_{N-1}) = \prod_{i=a}^b (1 + x_i + x_{i-1})$ and let

$$\langle h \rangle = \int_0^\infty dx_0 \cdots \int_0^\infty dx_{N-1} \exp\left(-\sum_{i=0}^{N-1} \alpha x_i\right) h(x_0, \ldots, x_{N-1}),$$

where $\alpha = 3 + \theta$. With this notation, we can write

$$Z_N(\alpha) = \hat{Z}_N(\theta) = \langle \Pi_0^{N-1} \rangle.$$

Removing the two terms involving x_0 from Π_0^{N-1}, we have

$$Z_N(\alpha) = \langle [(1+x_0)^2 + (1+x_0)x_1 + (1+x_0)x_{N-1} + x_1 x_{N-1}] \Pi_2^{N-1} \rangle. \tag{3.2}$$

Let $m_k = \int dx_0\, x_0^k e^{-\alpha x_0}$. Integrating by parts, we have $m_k = (k/\alpha) m_{k-1}$ and, hence,

$$m_0 = 1/\alpha, \qquad m_1 = 1/\alpha^2, \qquad m_2 = 2/\alpha^3, \qquad m_3 = 6/\alpha^4. \tag{3.3}$$

If we define

$$\begin{aligned} &Y_N = \langle \Pi_1^{N-1} \rangle, && W_N = \langle x_0\, \Pi_1^{N-1} \rangle && \text{for } N \ge 1, \\ &U_N = \langle \Pi_1^{N-1} x_{N-1} \rangle, && V_N = \langle x_0\, \Pi_1^{N-1} x_{N-1} \rangle && \text{for } N \ge 2, \end{aligned} \tag{3.4}$$

then we have, for $N \ge 3$,

$$Z_N(\alpha) = \left(\frac{1}{\alpha} + \frac{2}{\alpha^2} + \frac{2}{\alpha^3}\right) Y_{N-1} + \left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\right)(W_{N-1} + U_{N-1}) + \frac{1}{\alpha} V_{N-1} \tag{3.5}$$
$$= \left(\frac{\alpha^2 + 2\alpha + 2}{\alpha^3}\right) Y_{N-1} + 2\left(\frac{\alpha+1}{\alpha^2}\right) W_{N-1} + \frac{1}{\alpha} V_{N-1},$$

since symmetry dictates $W_{N-1} = U_{N-1}$.

To begin, we consider the pair Y_N, W_N. Since $\Pi_1^{N-1} = (1 + x_0)\Pi_2^{N-1} + x_1\Pi_2^{N-1}$ and $x_0\Pi_1^{N-1} = x_0(1 + x_0)\Pi_2^{N-1} + x_0 x_1 \Pi_2^{N-1}$, we have, for $N \ge 2$,

$$Y_N = \int_0^\infty dx_0\, e^{-\alpha x_0}(1 + x_0) Y_{N-1} + \int_0^\infty dx_0\, e^{-\alpha x_0} W_{N-1},$$
$$W_N = \int_0^\infty dx_0\, e^{-\alpha x_0} x_0 (1 + x_0) Y_{N-1} + \int_0^\infty dx_0\, e^{-\alpha x_0} x_0 W_{N-1}.$$

Using (3.3), we now have

$$Y_N = \left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\right) Y_{N-1} + \frac{1}{\alpha} W_{N-1}, \tag{3.6}$$
$$W_N = \left(\frac{1}{\alpha^2} + \frac{2}{\alpha^3}\right) Y_{N-1} + \frac{1}{\alpha^2} W_{N-1}.$$

Subtracting $1/\alpha$ times the first equation from the second, we get

$$W_N = \frac{1}{\alpha} Y_N + \frac{1}{\alpha^3} Y_{N-1}. \tag{3.7}$$

Substituting this into the first equation in (3.6), we have

$$Y_N = \left(\frac{1}{\alpha} + \frac{2}{\alpha^2}\right) Y_{N-1} + \frac{1}{\alpha^4} Y_{N-2}. \tag{3.8}$$

This is a second-order difference equation, $Y_N = aY_{N-1} + bY_{N-2}$, so its general solution will be of the form $C_1^Y \beta_1^N + C_2^Y \beta_2^N$, where β_1, β_2 are the two roots of $\beta^2 - a\beta - b$. In the special case $\alpha = 3$, we have $a = 5/9$ and $b = 1/81$. Using the quadratic formula, we find

$$\beta_1 = \frac{5 + \sqrt{29}}{18} = 0.5769536, \qquad \beta_2 = \frac{5 - \sqrt{29}}{18} = -0.021398.$$

Since $Y_N > 0$, we must have $C_1^Y > 0$, and it follows that

$$(1/N) \log Y_N \to \log \beta_1. \tag{3.9}$$

Formula (3.7) implies that

$$W_N = \left(\frac{1}{\alpha} + \frac{2}{\alpha^2}\right) W_{N-1} + \frac{1}{\alpha^4} W_{N-2}; \tag{3.10}$$

hence, $W_N = C_1^W \beta_1^N + C_2^W \beta_2^N$. Again, since $W_N > 0$, we must have $C_1^Z > 0$, and it follows that

$$(1/N)\log W_N \to \log \beta_1.$$

Repeating the reasoning above for the pair U_N, V_N, we find the same recursion,

$$U_N = \left(\frac{1}{\alpha} + \frac{1}{\alpha^2}\right) U_{N-1} + \frac{1}{\alpha} V_{N-1},$$
$$V_N = \left(\frac{1}{\alpha^2} + \frac{2}{\alpha^3}\right) U_{N-1} + \frac{1}{\alpha^2} V_{N-1},$$

and conclude that

$$U_N = \left(\frac{1}{\alpha} + \frac{2}{\alpha^2}\right) U_{N-1} + \frac{1}{\alpha^4} U_{N-2},$$
$$V_N = \left(\frac{1}{\alpha} + \frac{2}{\alpha^2}\right) V_{N-1} + \frac{1}{\alpha^4} V_{N-2} \tag{3.11}$$

and

$$(1/N)\log U_N \to \log \beta_1 \quad \text{and} \quad (1/N)\log V_N \to \log \beta_1. \tag{3.12}$$

From (3.5), (3.8), (3.10) and (3.11), we see that

$$Z_N(\alpha) = A(\alpha)\beta_1(\alpha)^N + B(\alpha)\beta_2(\alpha)^N,$$
$$\text{where } \beta_i(\alpha) = \frac{(1/\alpha + 2/\alpha^2) \pm \sqrt{(1/\alpha + 2/\alpha^2)^2 + 4/\alpha^4}}{2}. \tag{3.13}$$

Simplifying, we have

$$\beta_1(\alpha) = \frac{\alpha + 2 + \sqrt{\alpha^2 + 4\alpha + 8}}{2\alpha^2} \quad \text{and} \quad \beta_2(\alpha) = \frac{\alpha + 2 - \sqrt{\alpha^2 + 4\alpha + 8}}{2\alpha^2}. \tag{3.14}$$

Of course, when $\alpha = 3$ this reduces to $\beta_1 = (5 \pm \sqrt{29})/18$. Since $P(G_N) = Z_N(3)$, we have

$$\frac{1}{N}\log P(G_N) \to \log \beta_1,$$

so $\lambda_1 = \beta_1 = 0.5769536$.

Our next goal is to compute the height of all 0's given that it is a local maximum. To do this, we begin by noting

$$Z_N'(3) = -E\left(\sum_{i=0}^{N-1} x_i; G_N\right).$$

Differentiating (3.13), we have

$$Z_N'(\alpha) = A'(\alpha)\beta_1(\alpha)^N + A(\alpha)N\beta_1(\alpha)^{N-1}\beta_1'(\alpha)$$
$$+ B'(\alpha)\beta_2(\alpha)^N + B(\alpha)N\beta_2(\alpha)^{N-1}\beta_2'(\alpha). \tag{3.15}$$

Since $\beta_1(3) > \beta_2(3)$, combining this with (3.13) gives

$$\lim_{N\to\infty} \frac{1}{N} \frac{-Z_N'(3)}{Z_N(3)} = -\frac{\beta_1'(3)}{\beta_1(3)}.$$

Differentiating (3.13), we have

$$\begin{aligned} 2\beta_1'(\alpha) &= -\frac{1}{\alpha^2} - \frac{4}{\alpha^3} + \frac{1}{2}\left(\frac{1}{\alpha^2} + \frac{4}{\alpha^3} + \frac{8}{\alpha^4}\right)^{-1/2}\left(-\frac{2}{\alpha^3} - \frac{12}{\alpha^4} - \frac{32}{\alpha^5}\right) \\ &= -\frac{1}{\alpha^2} - \frac{4}{\alpha^3} + \frac{1}{2}(\alpha^2 + 4\alpha + 8)^{-1/2}\left(-\frac{2\alpha^2 - 12\alpha - 32}{\alpha^3}\right). \end{aligned}$$

Setting $\alpha = 3$, we have

$$2\beta'(3) = -\frac{7}{27} - \frac{86}{54\sqrt{29}}.$$

Since $\beta_1(3) = (5 + \sqrt{29})/18$, we have

$$-\frac{\beta_1'(3)}{\beta_1(3)} = \frac{7\sqrt{29} + 43}{54\sqrt{29}} \frac{18}{5 + \sqrt{29}} = \frac{126\sqrt{29} + 774}{270\sqrt{29} + 1566} = 0.480971328. \tag{3.16}$$

Note that although $E(x_i) = 1$ (since $x_i = -v_i$ and $Ev_i = -1$) we have

$$\lim_{N\to\infty} E(x_i|G_N) \approx 0.481.$$

To compute the variance, we note that

$$\begin{aligned} \operatorname{var}\left(\sum_{i=0}^{N-1} x_i \middle| G_N\right) &= E\left(\left(\sum_{i=0}^{N-1} x_i\right)^2 \middle| G_N\right) - \left\{E\left(\sum_{i=0}^{N-1} x_i \middle| G_N\right)\right\}^2 \\ &= \frac{Z_N''(3)}{Z_N(3)} - \left(\frac{Z_N'(3)}{Z_N(3)}\right)^2. \end{aligned} \tag{3.17}$$

Ignoring the terms involving $\beta_2(\alpha)$ that will vanish in the limit, we have

$$\begin{aligned} Z_N''(\alpha) &= A''(\alpha)\beta_1(\alpha)^N + 2A'(\alpha)N\beta_1(\alpha)^{N-1}\beta_1'(\alpha) \\ &\quad + A(\alpha)N(N-1)\beta_1(\alpha)^{N-2}\beta_1'(\alpha)^2 + A(\alpha)N\beta_1(\alpha)^{N-1}\beta_1''(\alpha). \end{aligned}$$

From this, it follows that

$$\begin{aligned} Z_N''(3) &= N^2 A(3)\beta_1(3)^{N-2}\beta_1'(3)^2 \\ &\quad + N\{2A'(3)\beta_1(3)^{N-1}\beta_1'(3) \\ &\qquad - A(3)\beta_1(3)^{N-2}\beta_1'(3)^2 + A(3)\beta_1(3)^{N-1}\beta_1''(3)\} \\ &\quad + O(\beta_2(3)^N) \end{aligned}$$

and, hence,

$$(3.18)\quad \frac{Z_N''(3)}{Z_N(3)} = N^2\left(\frac{\beta_1'(3)}{\beta_1(3)}\right)^2 + N\left\{2\frac{A'(3)}{A(3)}\frac{\beta_1'(3)}{\beta_1(3)} - \left(\frac{\beta_1'(3)}{\beta_1(3)}\right)^2 + \frac{\beta_1''(3)}{\beta_1(3)}\right\} + o(1).$$

Using (3.15) and ignoring the terms involving $\beta_2(\alpha)$, we have

$$\begin{aligned} Z_N'(3)^2 &= \bigl(NA(3)\beta_1(3)^{N-1}\beta_1'(3)\bigr)^2 + N\bigl(2A'(3)A(3)\beta_1(3)^{2N-1}\beta_1'(3)\bigr) \\ &\quad + O\bigl(\beta_1^N(3)\bigr), \\ Z_N(3)^2 &= \bigl(A(3)\beta_1(3)^N\bigr)^2 + O\bigl(\beta_1^N(3)\bigr) \end{aligned}$$

and, hence,

$$(3.19)\qquad \frac{Z_N'(3)^2}{Z_N(3)^2} = N^2\left(\frac{\beta_1'(3)}{\beta_1(3)}\right)^2 + N\left\{2\frac{A'(3)}{A(3)}\frac{\beta_1'(3)}{\beta_1(3)}\right\} + o(1).$$

Combining (3.17)–(3.19), we have

$$(3.20)\qquad \operatorname{var}\left(\sum_{i=0}^{N-1} x_i \,\middle|\, G_N\right) \sim N\left\{\frac{\beta_1''(3)}{\beta_1(3)} - \left(\frac{\beta_1'(3)}{\beta_1(3)}\right)^2\right\}.$$

Letting $\mu = -\beta_1'(3)/\beta_1(3)$ and using Chebyshev's inequality shows that, for any $\varepsilon > 0$,

$$(3.21)\qquad P\left(\left|\sum_{i=0}^{N-1} x_i - \mu N\right| > N^{1/2+\varepsilon} \,\middle|\, G_N\right) \to 0.$$

Using the analysis of $Z_N(\alpha)$, it is not hard to improve the last result to a central limit theorem. To do this, we begin by computing the Laplace transform

$$\begin{aligned} \psi(\theta) &\equiv E\left(\exp\left(-\frac{\theta}{\sqrt{N}}\left(\sum_{i=0}^{N-1} x_i - \mu N\right)\right); G_N\right) \\ &= \exp(\theta\mu\sqrt{N})Z_N\left(3 + \frac{\theta}{\sqrt{N}}\right). \end{aligned}$$

Since $\beta_1(\alpha) > \beta_2(\alpha)$ for α near 3, we have $Z_N(\alpha) \approx A(\alpha)\beta_1(\alpha)^N$ and, hence,

$$\frac{\psi(\theta)}{P(G_N)} \approx \exp(\theta\mu\sqrt{N})\frac{A(3+\theta/\sqrt{N})\beta_1(3+\theta/\sqrt{N})^N}{A(3)\beta_1(3)^N}.$$

As $N \to \infty$, we have $A(3+\theta/\sqrt{N})/A(3) \to 1$. Using Taylor's theorem with remainder, we see that the above

$$\sim \exp(\theta\mu\sqrt{N})\left(\frac{\beta_1(3) + (\theta/\sqrt{N})\beta_1'(3) + (\theta^2/2N)\beta_1''(\alpha_N)}{\beta_1(3)}\right)^N,$$

where $\alpha_N \in (3, 3+\theta/\sqrt{N})$. Taking logarithms, we find

$$\log\bigl(\psi(\theta)/P(G_N)\bigr) = \theta\mu\sqrt{N} + N\log\biggl(1 + \frac{\theta}{\sqrt{N}}\frac{\beta_1'(3)}{\beta_1(3)} + \frac{\theta^2}{2N}\frac{\beta_1''(\alpha_N)}{\beta_1(3)}\biggr).$$

Using $\log(1+x) = x - x^2/2 + \cdots$, we have that the right-hand side

$$\sim \frac{\theta^2}{2}\biggl[-\biggl(\frac{\beta_1'(3)}{\beta_1(3)}\biggr)^2 + \frac{\beta_1''(3)}{\beta_1(3)}\biggr].$$

Letting σ^2 denote the term in square brackets and recalling that the Laplace transform of the normal with mean 0 and variance σ^2 is $\exp(\sigma^2\theta^2/2)$, we have shown the following result.

THEOREM 3.1. *As* $N \to \infty$,

$$P\Biggl(\Biggl(\sum_{i=0}^{N-1} x_i - \mu N\Biggr)/\sigma\sqrt{N} \le x\Bigg|G_N\Biggr) \to P(\chi \le x),$$

where χ *has a normal distribution with mean* 0 *and variance* 1.

The last detail is to compute σ^2. To get $\beta_1''(3)$, we differentiate (3.13) twice to get

$$\begin{aligned}
2\beta_1''(\alpha) &= \frac{2}{\alpha^3} + \frac{12}{\alpha^4} - \frac{1}{4}\biggl(\frac{1}{\alpha^2} + \frac{4}{\alpha^3} + \frac{8}{\alpha^4}\biggr)^{-3/2}\biggl(-\frac{2}{\alpha^3} - \frac{12}{\alpha^4} - \frac{32}{\alpha^5}\biggr)^2 \\
&\quad + \frac{1}{2}\biggl(\frac{1}{\alpha^2} + \frac{4}{\alpha^3} + \frac{8}{\alpha^4}\biggr)^{-1/2}\biggl(\frac{6}{\alpha^4} + \frac{48}{\alpha^5} + \frac{160}{\alpha^6}\biggr) \\
&= \frac{2}{\alpha^3} + \frac{12}{\alpha^4} - \frac{1}{4}(\alpha^2 + 4\alpha + 8)^{-3/2}\biggl(\frac{2\alpha^2 - 12\alpha - 32}{\alpha^2}\biggr)^2 \\
&\quad + \frac{1}{2}(\alpha^2 + 4\alpha + 8)^{-1/2}\biggl(\frac{6\alpha^2 + 48\alpha + 160}{\alpha^4}\biggr).
\end{aligned}$$

Setting $\alpha = 3$, we have

$$\begin{aligned}
2\beta_1''(3) &= \frac{6}{27} - \frac{1}{4\cdot 29\cdot\sqrt{29}}\biggl(\frac{86}{9}\biggr)^2 + \frac{1}{2\sqrt{29}}\cdot\frac{358}{27} \\
&= \frac{18\cdot 29\cdot\sqrt{29} - 43^2 + 179\cdot 3\cdot 29}{81\cdot 29\cdot\sqrt{29}} \\
&= \frac{522\sqrt{29} + 13724}{2349\sqrt{29}}.
\end{aligned}$$

Dividing by 2 and by $\beta_1(3) = (5+\sqrt{29})/18$, we have

$$\frac{\beta_1''(3)}{\beta_1(3)} = \frac{261\sqrt{29}+6862}{2349\sqrt{29}} \frac{18}{5+\sqrt{29}} = \frac{4698\sqrt{29}+123516}{11745\sqrt{29}+68121} = 1.132799243.$$

From this, it follows that $\sigma^2 = 0.901465824$.

Theorem 3.1 gives the approximate distribution of the height of a local maximum chosen at random. Our next result considers the height of the global maximum, H_N^*.

THEOREM 3.2. *Let* $b = \inf\{a \in (-1,0) : -2a\exp(a+1) < 1\} \approx -0.231961$. *If* $a > b$, *then* $P(H_N^* > aN) \to 0$.

PROOF. Let $h(\theta) = \int_{-\infty}^0 e^{\theta x} e^x \, dx = 1/(1+\theta)$ be the Laplace transform of the negative exponential. If S_N is the sum of N independent random variables with a negative exponential distribution, then Markov's inequality implies that if $\theta > 0$ then

$$e^{\theta N a} P(S_N > Na) \le h(\theta)^N. \tag{3.22}$$

Rearranging, we have

$$P(S_N > Na) \le \exp\bigl(-N[\theta a + \log(1+\theta)]\bigr). \tag{3.23}$$

To optimize the estimate, we set

$$0 = \frac{d}{d\theta}[\theta a + \log(1+\theta)] = a + \frac{1}{1+\theta}. \tag{3.24}$$

When $a \in (-1,0)$, the solution is $\theta = -(1+a)/a > 0$, so we have

$$P(S_N > Na) \le \exp\bigl(N[a+1-\log(-1/a)]\bigr).$$

Since there are 2^N possible sequences (proteins) and to each corresponds a sum of N independent negative exponentials, we have

$$P(H_N^* > aN) \le (-2a)^N \exp\bigl(N(a+1)\bigr). \tag{3.25}$$

Taking $b = \inf\{a \in (-1,0) : -2a\exp(a+1) < 1\}$, the desired result follows. Since $a \to a+1-\log(-1/a)$ is increasing, it is straightforward to compute numerically that $b = -0.231961$. The reader can verify the computation by checking that $-2b\exp(b+1) \approx 1$. $\square$

Using the same technique, one can obtain the following result.

THEOREM 3.3. *Consider the positive exponential distribution and let* $b = \{a > 1 : 2a\exp(1-a) < 1\} \approx 2.678347$. *If* $a > b$, *then* $P(H_N^* > aN) \to 0$.

PROOF. Let $h(\theta) = \int_0^\infty e^{\theta x} e^{-x}\, dx = 1/(1-\theta)$ be the Laplace transform of the (positive) exponential (rate 1). Let S_N be the sum of N independent exponential (rate 1) random variables. The analogues of (3.22)–(3.24) are now

$$e^{\theta N a} P(S_N > Na) \le h(\theta)^N,$$
$$P(S_N > Na) \le \exp\bigl(-N[\theta a + \log(1-\theta)]\bigr)$$

and

$$0 = \frac{d}{d\theta}[\theta a + \log(1-\theta)] = a - \frac{1}{1-\theta}.$$

When $a > 1$, the solution $\theta = 1 - 1/a > 0$, so we have

$$P(S_N > Na) \le \exp\bigl(-N[a - 1 + \log(1/a)]\bigr).$$

As in (3.25), we get

$$P(H_N^* > aN) \le (2a)^N \exp\bigl(-N(a-1)\bigr),$$

and, after taking $b = \inf\{a > 1 : 2a \exp(1-a) < 1\}$, the desired result follows. □

4. Results for the uniform case, $K = 1$. When $K = 1$ and $f(y) = 1$ for $0 \le y \le 1$, we have

$$p(x, y) = F_2(x + y), \qquad x, y \in [0, 1]. \tag{4.1}$$

Since this is symmetric and square integrable, a theorem on page 243 of Riesz and Nagy (1990) implies that we can write

$$p(x, y) = \sum_{i=1}^{\infty} \beta_i h_i(x) h_i(y), \tag{4.2}$$

where β_i is a decreasing sequence of eigenvalues and the $h_i(x)$ are the corresponding eigenfunctions, which form an orthonormal sequence.

Iterating and using the fact that the h_i are orthonormal, we have

$$p^2(x, y) = \int p(x, u) p(u, y)\, du = \sum_{i=1}^{\infty} \beta_i^2 h_i(x) h_i(y)$$

or, in general, that

$$p^n(x, y) = \sum_{i=1}^{\infty} \beta_i^n h_i(x) h_i(y). \tag{4.3}$$

In Section 6, we will show that $Rp(x, y)h(y)/h(x)$ is a transition kernel of a Harris chain with unique stationary distribution, so $\beta_1 > \beta_2$ and if $R = 1/\beta_1$ we have

$$R^n p^n(x, y) \to h_1(x) h_1(y). \tag{4.4}$$

As in (2.12), we have

$$P(G'_{N-1}) = \int dx_0 \int dx_{N-1}\, p^{N-1}(x_0, x_{N-1}),$$

and it follows that

$$R^{N-1} P(G'_{N-1}) \to \left(\int h_1(x)\, dx \right)^2. \tag{4.5}$$

Comparing with (2.2), we see that $\lambda_1 = \beta_1 = 1/R$.

Combining this result with the variational characterization of the largest eigenvalue

$$\beta_1 = \max \frac{\iint g(x)p(x, y)g(y)\, dx\, dy}{\int g(x)^2\, dx}$$

allows us to get a lower bound on λ_1. A little calculus shows that

$$p(x, y) = \begin{cases} (x+y)^2/2, & 0 \le x+y \le 1, \\ 1-(2-x-y)^2/2, & 1 \le x+y \le 2, \end{cases}$$

and that

$$\iint p(x, y)\, dx\, dy = \tfrac{1}{2},$$

$$\iint xp(x, y)\, dx\, dy = \tfrac{37}{120},$$

$$\iint xyp(x, y)\, dx\, dy = \tfrac{11}{60}.$$

Taking $g(x) = 1 + ax$, we have

$$\lambda_1 \ge \frac{\frac{1}{2} + \frac{37}{60}a + \frac{11}{60}a^2}{1 + a + a^2/3}.$$

Differentiating the right-hand side with respect to a gives

$$\frac{(\frac{37}{60} + \frac{22}{60}a)(1 + a + a^2/3) - (\frac{1}{2} + \frac{37}{60}a + \frac{11}{60}a^2)(1 + 2a/3)}{(1 + a + a^2/3)^2}.$$

Fortunately, the cubic terms cancel out, and 180 times the numerator becomes

$$(111 - 90) + a(111 + 66 - 111 - 60) + a^2(37 + 66 - 33 - 74) = 21 + 6a - 4a^2.$$

Solving the quadratic equation, we have $a = 3.1609127$ and

$$\lambda_1 \ge 0.571455. \tag{4.6}$$

It is important that the lower bound is greater than $1/2$, since the expected number of local maxima is $EM_N = 2^N P(G_N)$ and we have $\lim_{N\to\infty}(1/N)\log EM_N = \log(2\lambda_1) > 0$.

To get a bound in the other direction, we use (2.11), which says that

$$\frac{1}{N}\log q_N^* \to \inf_{M\ge 1}\frac{1}{M}\log q_M^* = \log\lambda_1,$$

and (2.10), which implies that $q_n^* = q_n(1)$. Mathematica computes (after three days of calculation) that $q_5^* = 0.0839578$, implying

$$\lambda_1 \le (q_5^*)^{1/5} = 0.609273. \tag{4.7}$$

5. Bounds on $\boldsymbol{\lambda_K}$ for large $\boldsymbol{K}$. In this section, we will derive lower bounds for $P(G_N)$, which show that $\lambda_K \to 1$ as $K\to\infty$. We begin with the situation in which F has a standard normal distribution. Due to the Cauchy–Schwarz inequality, we have

$$EW \ge (EW^{1/2}V)^2/EV^2. \tag{5.1}$$

We will apply (5.1) with $W = \prod_{i=0}^{N-1} F_{K+1}(X_{i-K}+\cdots+X_i)$ and

$$V = \frac{dQ}{dP} = \frac{\exp(\xi\sum_{i=0}^{N-1}X_i)}{\psi(\xi)^N},$$

where $\psi(\theta)$ is the moment-generating function

$$\psi(\theta) = \int \frac{1}{\sqrt{2\pi}}\exp\biggl(-\frac{x^2}{2}\biggr)\exp(\theta x)\,dx = \exp\biggl(\frac{\theta^2}{2}\biggr).$$

Note that if, under P, the coordinates of $(X_0,\dots,X_{N-1})$ are i.i.d. normal variables with mean 0 and variance 1, then, under Q, the coordinates of $(X_0,\dots,X_{N-1})$ are independent and identically distributed normals with mean ξ and variance 1. The change of measure and Harris's inequality imply that

$$E\biggl(W^{1/2}\frac{dQ}{dP}\biggr) = E_Q(W^{1/2}) \ge \bigl[E_Q\bigl(F_{K+1}^{1/2}(X_0+\cdots+X_K)\bigr)\bigr]^N. \tag{5.2}$$

We choose

$$\xi = 2\sqrt{\frac{2\log(K+1)}{K+1}},$$

so that the expected value under Q, $E_Q(F_{K+1}^{1/2}(X_0+\cdots+X_K))$, converges to 1 as $K\to\infty$.

To compute the right-hand side of (5.2), we note that, under Q, $X_0+\cdots+X_K$ is normal with mean $\mu_K = 2\sqrt{2(K+1)\log(K+1)}$ and variance $K+1$. Scaling to express things in terms of a standard normal random variable χ, we get

$$\begin{aligned}
&E_Q\bigl(F_{K+1}^{1/2}(X_0+\cdots+X_K)\bigr)\\
&\quad\ge Q(X_0+\cdots+X_K\ge \mu_K/2)F_{K+1}^{1/2}(\mu_K/2)\\
&\quad= P\bigl(\chi \le \sqrt{2\log(K+1)}\bigr)^{3/2}.
\end{aligned}$$

Using the fact that $P(\chi > x) \le \exp(-x^2/2)$ in the last inequality, we have

$$E_Q\big(F_{K+1}(X_0 + \cdots + X_K)^{1/2}\big) \ge \left(1 - \frac{1}{K+1}\right)^{3/2}. \tag{5.3}$$

It remains to bound the second moment of the Radon–Nikodym derivative from above. Note that

$$\begin{aligned} E\left(\frac{dQ}{dP}\right)^2 &= \frac{E\exp(2\xi\sum_{i=0}^{N-1} X_i)}{\psi(\xi)^{2N}} \\ &= \exp(\xi^2 N) = \exp\left(\frac{8\log(K+1)}{K+1}N\right). \end{aligned} \tag{5.4}$$

Combining (5.1)–(5.4), we have

$$EW \ge \left(1 - \frac{1}{K+1}\right)^{3N/2} \exp\left(-\frac{8\log(K+1)}{K+1}N\right),$$

implying

$$\begin{aligned} \lambda_K &\ge \left(1 - \frac{1}{K+1}\right)^3 \exp\left(-\frac{8\log(K+1)}{K+1}\right) \\ &\ge 1 - \frac{9\log(K+1)}{K+1} \end{aligned} \tag{5.5}$$

for large K.

The derivation of the last result only involved values of the moment-generating function when θ was close to 0, so it will hold for distributions where the moment-generating function is finite in a neighborhood of 0. Without loss of generality, we can suppose that the mean and the variance of F satisfy $\mu = 0$ and $\sigma^2 = 1$. Inequalities (5.1) and (5.2) hold in general, so we begin with the estimation of the right-hand side of (5.2). Due to our assumptions, we have $\psi'(0) = 0$ and $\psi''(0) = 1/2$, where $\psi(\theta) = \int e^{\theta x} dF(x)$. Therefore, if $|\theta|$ is sufficiently small, we have

$$\exp\left(\frac{\theta^2}{2.02}\right) \le 1 + \frac{\theta^2}{2.01} \le \psi(\theta) \le 1 + \frac{\theta^2}{1.99} \le \exp\left(\frac{\theta^2}{1.99}\right).$$

Let

$$\nu_K = \sqrt{2.1\log(K+1)/(K+1)}.$$

Markov's inequality implies that, if $\theta > 0$ is small,

$$\exp(\theta\nu_K(K+1))\big[1 - F_{K+1}\big(\nu_K(K+1)\big)\big] \le \exp\left(\frac{\theta^2(K+1)}{1.99}\right).$$

Taking $\theta = \nu_K$, we have

$$1 - F_{K+1}\big(\nu_K(K+1)\big) \le \exp\big(-\nu_K^2(K+1)(0.99)/1.99\big) \le (K+1)^{-1}. \tag{5.6}$$

To bound $Q(X_0+\cdots+X_K \ge \nu_K(K+1))$, we note that, under Q, X_i has moment-generating function $\psi(\theta+\xi)/\psi(\xi)$, so

$$\exp\bigl(\theta\nu_K(K+1)\bigr)Q\bigl(X_0+\cdots+X_K \le \nu_K(K+1)\bigr) \le \frac{\exp((\xi+\theta)^2(K+1)/1.99)}{\exp(\xi^2(K+1)/2.02)}.$$

Taking $\theta = \nu_K - \xi$, where $|\xi| \le 5\nu_K$, we have

$$\begin{aligned} &Q\bigl(X_0+\cdots+X_K \le \nu_K(K+1)\bigr) \\ &\quad\le \exp\biggl(-(K+1)\biggl\{\frac{\xi^2}{2.02} - (\xi-\nu_K)\nu_K - \frac{\nu_K^2}{1.99}\biggr\}\biggr). \end{aligned}$$

Setting $\xi = 2.01\nu_K$, the above inequality becomes

$$\begin{aligned} &\le \exp\biggl(-(K+1)\nu_K^2\biggl\{\frac{4.0401}{2.02} - 1.01 - \frac{1}{1.99}\biggr\}\biggr) \\ &\le \exp\bigl(-(2.1)(0.487)\log(K+1)\bigr) \le \frac{1}{K+1}. \end{aligned} \tag{5.7}$$

Combining (5.6) and (5.7), we have

$$\begin{aligned} &E_Q\bigl(F_{K+1}(X_0+\cdots+X_K)^{1/2}\bigr) \\ &\quad\ge Q\bigl(X_0+\cdots+X_K \ge \nu_K(K+1)\bigr)F_{K+1}^{1/2}\bigl(\mu_K(K+1)\bigr) \\ &\quad\ge \biggl(1-\frac{1}{K+1}\biggr)^{3/2}. \end{aligned} \tag{5.8}$$

Similarly, note that

$$\begin{aligned} E\biggl(\frac{dQ}{dP}\biggr)^2 &= \frac{\psi(2\xi)^N}{\psi(\xi)^{2N}} \\ &\le \exp\biggl(N\xi^2\biggl(\frac{4}{1.99} - \frac{2}{2.02}\biggr)\biggr) \\ &= \exp\biggl(N(2.01)^2\frac{2.1\log(K+1)}{K+1}\frac{8.08-3.98}{(2.02)(1.99)}\biggr) \\ &\le \exp\biggl(8.7\frac{\log(K+1)}{K+1}N\biggr). \end{aligned}$$

Using the last result together with (5.8), (5.2) and (5.1) we have, as in (5.5),

$$\lambda_K \ge \biggl(1-\frac{1}{K+1}\biggr)^{3/2}\exp\biggl(-\frac{8.7\log(K+1)}{K+1}\biggr) \ge 1 - \frac{9\log(K+1)}{K+1}$$

for large K.

6. Results for the height of local maxima, $K \geq 1$. We will use the R-theory of Markov chains as developed by Tweedie (1974). Let $R = \sup\{r : r^n q^n(x, y) \to 0\}$. Our first goal is to check the R-positive recurrence condition given on page 844 of Tweedie (1974) for the transition probability

$$q\big((x_0, \dots, x_{K-1}), (x_1, \dots, x_K)\big) = F_{K+1}(x_0 + \cdots + x_K) f(x_K).$$

Let $\Lambda = 1/\lambda_K$. It follows from (2.11) that $R \geq \Lambda$. To prove that $\Lambda \geq R$ and the chain is R-recurrent, it suffices to show that $\Lambda^n q^n(x, y) \not\to 0$ for some fixed x and y. Note that (2.11) implies

$$\Lambda^n q_n^* \geq 1, \qquad n \geq 1. \tag{6.1}$$

Let a be in the interior of the support of f. Suppose first that $K = 1$. Using the Markov property,

$$\Lambda^n q^n(x, a) = \int \Lambda^{n-1} q^{n-1}(x, y) \Lambda q(y, a)\, dy.$$

Since the integrand above is nonnegative and $q(y, a) = F_2(y + a) f(a) \geq F_2(b + a) f(a)$ for $y \geq b$, we have

$$\Lambda^n q^n(x, a) \geq \int_b^\infty \Lambda^{n-1} q^{n-1}(x, y)\, dy\, \Lambda F_2(b + a) f(a). \tag{6.2}$$

To estimate $\int_{-\infty}^b \Lambda^{n-1} q^{n-1}(x, y)\, dy$, we use again the Markov property

$$q^{n-1}(x, y) = \int q^{n-2}(x, z) q(z, y)\, dz,$$

and note that

$$\int_{-\infty}^b q(z, y)\, dy = \int_{-\infty}^b F_2(z + y) f(y)\, dy \leq F_2(z + b) \int_{-\infty}^b f(y)\, dy,$$
$$\int_b^\infty q(z, y)\, dy \geq F_2(z + b) \int_b^\infty f(y)\, dy.$$

Therefore, if we pick b so that $\int_{-\infty}^b f(y)\, dy = \int_b^\infty f(y)\, dy = 1/2$, that is, b is a median of F, then we have

$$\int_b^\infty q(z, y)\, dy \geq \tfrac{1}{2} \int_{-\infty}^\infty q(z, y)\, dy,$$

and, by the sentence following (6.2) and Fubini's theorem,

$$\int_b^\infty \Lambda^{n-1} q^{n-1}(x, y)\, dy \geq \tfrac{1}{2} \int \Lambda^{n-1} q^{n-1}(x, y)\, dy. \tag{6.3}$$

Taking $\sup_x$ and using (6.1) and (6.2), we have

$$\sup_x \Lambda^n q^n(x, a) \geq \frac{\Lambda}{2} F_2(b + a) f(a) > 0.$$

At this point, we consider two cases:

CASE 1. $F(x) = 1$ for some $x < \infty$ and, without loss of generality, $1 = \inf\{x : F(x) = 1\}$. Then

$$\Lambda^n q^n(1, a) \geq \frac{\Lambda}{2} F_2(b + a) f(a).$$

CASE 2. If $F(x) < 1$ for all x, then

$$\sup_x \Lambda^n q^n(x, a) = \Lambda^n \int dy_0\, f(y_0) q^{n-1}(y_0, a).$$

Using the Markov property and the monotonicity of F_2, we get

$$\begin{aligned} \sup_x \Lambda^n q^n(x, a) &= \int dy_0\, f(y_0) \int dy_1\, F_2(y_0 + y_1) f(y_1) \Lambda^n q^{n-2}(y_1, a) \\ &\leq 2 \int_b^\infty dy_0\, f(y_0) \Lambda^n q^{n-1}(y_0, a). \end{aligned} \tag{6.4}$$

The last inequality is a consequence of calculations similar to those that led to (6.3). Combining (6.4) with the observation that

$$\Lambda^n q^n(a, a) \geq F(a + b) \int_b^\infty dy_0\, f(y_0) \Lambda^n q^{n-1}(y_0, a), \tag{6.5}$$

we have

$$\Lambda^n q^n(a, a) \geq \frac{F(a + b)}{2} \sup_x \Lambda^n q^n(x, a),$$

and the desired result follows.

The above argument generalizes to $K \geq 1$ in the following way. Let $W_1^+, \dots, W_K^+$ be i.i.d. with density function $2 f(x) \mathbb{1}_{\{x > b\}}$ and let $W_1^-, \dots, W_K^-$ be i.i.d. with density function $2 f(x) \mathbb{1}_{\{x < b\}}$. For each subset $I \subset \{1, 2, \dots, K\}$, define $A_I = \{x : x_i > b$ if and only if $i \in I\}$. For each $i \in \{1, \dots, K\}$, let

$$Y_i^I = \begin{cases} W_i^+, & i \in I, \\ W_i^-, & i \notin I, \end{cases} \quad \text{and} \quad Z_i = W_i^+.$$

Clearly, $Y_i^I \leq Z_i$ for all I, i and, due to the monotonicity of F_{K+1},

$$\begin{aligned} 2^K \int_{A_I} q^K(x, y)\, dy &= E\left(\prod_{j=1}^K F_{K+1}(x_j + \cdots + x_K + Y_i^I + \cdots + Y_j^I) \right) \\ &\leq E\left(\prod_{j=1}^K F_{K+1}(x_j + \cdots + x_K + Z_1 + \cdots + Z_j) \right) \\ &= 2^K \int_{[b,\infty)^K} q^K(x, y)\, dy. \end{aligned}$$

Since $\mathbf{R}^K$ is a disjoint union of A_I over all subsets I of $\{1, \dots, K\}$, we get

$$\int q^K(x, y)\, dy \le 2^K \int_{[b,\infty)^K} q^K(x, y)\, dy.$$

Now, if $n \ge K$,

$$\begin{aligned}\int q^n(w, y)\, dy &= \int\int q^{n-K}(w, x) q^K(x, y)\, dy\, dx \\ &\le 2^K \int\int_{[b,\infty)^K} q^{n-K}(w, x) q^K(x, y)\, dy\, dx \\ &= 2^K \int_{[b,\infty)^K} q^n(w, y)\, dy,\end{aligned}$$

so that due to (6.1), for each $n \ge K$,

$$\sup_w \int_{[b,\infty)^K} \Lambda^n q^n(w, y)\, dy \ge 2^{-K}.$$

Let $\bar{a} = (a, \dots, a) \in \mathbf{R}^K$. As in the one-dimensional case, monotonicity implies

$$\begin{aligned}\Lambda^n q^n(x, \bar{a}) &\ge (\Lambda f(a))^K \prod_{i=1}^K F_{K+1}\big(bi + (K + 1 - i)a\big) \\ &\quad \times \int_{[b,\infty)^K} \Lambda^{n-K} q^{n-K}(x, y)\, dy,\end{aligned} \tag{6.6}$$

so that, for $n \ge 2K$,

$$\sup_x \Lambda^n q^n(x, \bar{a}) \ge (\Lambda f(a))^K \prod_{i=1}^K F_{K+1}\big(bi + (K + 1 - i)a\big) 2^{-K} > 0.$$

We again have two cases. If $F(x) = 1$ for some $x < 1$, then clearly for this x we have

$$\Lambda^n q^n(\bar{x}, \bar{a}) \ge (\Lambda f(a))^K \prod_{i=1}^K F_{K+1}\big(bi + (K + 1 - i)a\big) 2^{-K},$$

where $\bar{x} = (x, \dots, x) \in \mathbf{R}^K$. If $F(x) < 1$ for all $x < \infty$, then instead of (6.4) we have (again by using $\mathbf{R}^K = \bigcup_I A_I$ and the random variables Y_i^I, Z_i)

$$\sup_x \Lambda^n q^n(x, \bar{a}) \le 2^K \int_{[b,\infty)^K} dy_0 \cdots dy_{K-1}\, f(y_0) \cdots f(y_{K-1}) \Lambda^n q^{n-K}(y, a),$$

and instead of (6.5), we have

$$\begin{aligned}\Lambda^n q^n(\bar{a}, \bar{a}) &\ge \prod_{i=1}^K F_{K+1}\big(bi + (K + 1 - i)a\big) \\ &\quad \times \int_{[b,\infty)^K} dy_0 \cdots dy_{K-1}\, f(y_0) \cdots f(y_{K-1}) \Lambda^n q^{n-K}(y, \bar{a}).\end{aligned}$$

At this point, we have shown that the chain with transition probability q is R-recurrent in the sense of Tweedie (1974). Let

$$Q(y, A) = \int_A q(y, z)\, dz.$$

Results in Section 3 of Tweedie's paper now imply the existence of a σ-finite measure μ and a nonnegative function h unique up to constant multiples so that

$$\int \mu(dy) R Q(y, A) = \mu(A) \quad \text{and} \quad \int R Q(y, dz) h(z) = h(y). \tag{6.7}$$

LEMMA 6.1. *The measure μ has a density $j(y)$ with respect to the Lebesgue measure.*

PROOF. The kernel $Q^K(y, dz)$ has density $q^K(y, z)$. Using

$$\mu(A) = \int \mu(dy) R^K \int_A q^K(y, z)\, dz = \int_A dz \int \mu(dy) R^K q^K(y, z),$$

we conclude $\mu(dz) = j(z)\, dz$, where $j(z) = \int \mu(dy) R^K q^K(y, z)$. $\square$

LEMMA 6.2. *For $y = (y_0, \dots, y_{K-1})$, let $\hat{y} = (y_{K-1}, \dots, y_0)$ and $g(y) = f(y_0) \cdots f(y_{K-1})$. There are constants $C_i \in (0, \infty), i = 1, 2$, so that $h(y) = C_1 j(\hat{y})/g(\hat{y})$ and $j(y) \le C_2 g(y)$. The measure μ is finite and the function h is bounded above.*

PROOF. Writing dy or $d\hat{y}$ as shorthand for $dy_0 \cdots dy_{K-1}$, we have

$$\begin{aligned}
\frac{j(\hat{z})}{g(\hat{z})} &= \frac{1}{g(\hat{z})} \int j(\hat{y}) R^K q^K(\hat{y}, \hat{z})\, d\hat{y} \\
&= \int g(\hat{y}) R^K q^K(\hat{y}, \hat{z}) g(\hat{z})^{-1} \frac{j(\hat{y})}{g(\hat{y})}\, d\hat{y} \\
&= \int R^K q^K(z, y) \frac{j(\hat{y})}{g(\hat{y})}\, dy,
\end{aligned}$$

where the last equality follows from (2.7). The uniqueness in (6.7) now implies that $h(y) = C_1 j(\hat{y})/g(\hat{y})$.

Next, we show that μ in (6.7) is a finite measure. Take a constant $a > 0$ and let $A = \{x : h(x) \ge a\}$. Clearly, $g(\hat{y}) = g(y)$. Then

$$\mu(A^c) = \int \frac{1}{C_1} h(x) g(x) \mathbb{1}_{A^c}\, dx \le \frac{a}{C_1} \int g(x)\, dx < \infty.$$

Also,

$$\mu(A) = \int j(x) \mathbb{1}_A\, dx \le \frac{1}{a} \int h(x) \mathbb{1}_A(x) j(x)\, dx \le \frac{1}{a} \int h(x) j(x)\, dx = \int h\, d\mu,$$

which is finite by Theorem 7(ii) in Tweedie (1974) and R-positivity.

We will now use the fact that μ is a finite measure to verify the inequality and boundedness of h. Note that $q^K(y,z) \le g(z)$, so

$$j(z) = \int j(y) R^K q^K(y,z)\,dy \le R^K \int j(y)\,dy\,g(z).$$

Finally, note that the above inequality says

$$h(\hat{z}) = \frac{C_1 j(z)}{g(z)} \le R^K \int j(x)\,dx$$

so h is bounded. □

Theorem 6 on page 860 of Tweedie (1974) implies that

$$R^n Q^n(x, A) \to \frac{\mu(A)h(x)}{\int \mu(dy)h(y)}. \tag{6.8}$$

At first, it may look like the eigenfunctions are in the wrong places. To check this formula, recall that if there were no killing then $R = 1$ and the right eigenfunction would identically equal 1. To simplify (6.8), we will now let $\pi(y) = cj(y)$, where the constant is chosen so that $\int dy\,\pi(y)h(y) = 1$.

To get information about $P(G_N)$ from this, note that

$$\begin{aligned} R^{N-K} P(G_N) = \int \cdots \int dx_0 \cdots dx_{K-1}\,dx_{N-K} \cdots dx_{N-1}\,R^{N-K} \\ \times q^{N-K}\big((x_0,\ldots,x_{K-1}),(x_{N-K},\ldots,x_{N-1})\big) \\ \times \prod_{i=0}^{K-1} F_{K+1}(x_{i-K},\ldots,x_i) f(x_i). \end{aligned} \tag{6.9}$$

Letting $N \to \infty$ and using (6.8), we see that the above converges to

$$\begin{aligned} \int \cdots \int dx_{-K} \cdots dx_{-1}\,dx_0 \cdots dx_K \\ \times \pi(x_{-K},\ldots,x_{-1}) h(x_0,\ldots,x_{K-1}) \prod_{i=0}^{K-1} F_{K+1}(x_{i-K},\ldots,x_i) f(x_i), \end{aligned}$$

which is finite by Lemma 6.2. The last computation implies that

$$P(G_N) \sim c/R^N, \tag{6.10}$$

sharpening the conclusion in Theorem 2.1.

To investigate the properties of coordinates of local maxima, it is useful to introduce the transformed chain

$$\bar{q}(x,y) = \frac{R}{h(x)} q(x,y) h(y). \tag{6.11}$$

Relation (6.7) and the irreducibility of Q imply that $h(x_1, \dots, x_k)$ is positive $\prod_{i=1}^{K} f(x_i)\, dx_i$ almost everywhere, so there are no problems caused by dividing by 0. Since $h(y)$ is a right eigenvector, the new kernel has $\int \bar{q}(x, y)\, dy = 1$. Since $\pi(x)$ is a left eigenvector, $\bar{\pi}(x) = \pi(x)h(x)$ is a stationary distribution

$$\int dx\, \pi(x)h(x)\bar{q}(x, y) = \pi(y)h(y).$$

It is easy to see that $\bar{q}(x, y)$ is a Harris chain.

Let P_N be the distribution of $(x_0, \dots, x_{N-1})$ conditioned on G_N. Let Q_N be the distribution of $(x_0, \dots, x_{N-1})$ under the Markov chain with transition probability $\bar{q}$ and initial distribution $\bar{\pi}$. From (2.1), the display following (2.5), (6.10) and (6.11), we see that the Radon–Nikodym derivative of P_N relative to Q_N may be written as follows:

$$\text{(6.12)}\quad \frac{dP_N}{dQ_N} \sim C \frac{g(x_0, \dots, x_{K-1})}{\pi(x_0, \dots, x_{K-1})} \prod_{i=0}^{K-1} F_{K+1}(x_{i-K}, \dots, x_i) \frac{1}{h(x_{N-K}, \dots, x_{N-1})}.$$

As we will see in a moment, standard results for Harris chains give us results under Q_N. To transfer these to P_N, we will use the following result.

LEMMA 6.3. *Given $c > -\infty$, there are constants $C_{3,c}, C_{4,c} < \infty$ so that $g(z)/\pi(z) \le C_{3,c}$ and $1/h(z) \le C_{4,c}$ when $z_j \ge c$ for $0 \le j \le K - 1$.*

PROOF. The reasoning that led to (6.6) implies that if $z_j \ge c$ and $0 \le j \le K - 1$, then

$$\sup_x \Lambda^n q^n(x, z) \ge \prod_{j=1}^{K} \bigl(\Lambda f(z_{n-j})\bigr) F_{K+1}\bigl(bi + (K + 1 - i)c\bigr)2^{-K}.$$

Again, there are two cases to consider as in the proof of R-recurrence.

CASE 1. $F(x) = 1$ for some $x < \infty$ and, without loss of generality, $1 = \inf\{x : F(x) = 1\}$. Then

$$\Lambda^n q^n(\bar{1}, z) \ge \prod_{j=1}^{K} \bigl(\Lambda f(z_{n-j})\bigr) F_{K+1}\bigl(bi + (K + 1 - i)c\bigr)2^{-K}.$$

Letting $n \to \infty$ and using (6.8) with $R = \Lambda$ and A a small ball centered at z, we get

$$\frac{\pi(z_0, \dots, z_{K-1})}{g(z_0, \dots, z_{K-1})} \ge \left(\frac{\Lambda}{2}\right)^K \frac{1}{h(\bar{1})} \prod_{j=1}^{K} F_{K+1}\bigl(bi + (K + 1 - i)c\bigr).$$

CASE 2. If $F(x) < 1$ for all x, then, similarly,

$$R^K \int_{\mathbf{R}^K} dy_0 \cdots dy_{K-1}\, f(y_0)\cdots f(y_{K-1}) h(y_0, y_1, \dots, y_{K-1}) \pi(z_0, \dots, z_{K-1})$$
$$= \lim_n \int_{\mathbf{R}^K} dy_0 \cdots dy_{K-1}\, f(y_0)\cdots f(y_{K-1}) R^n q^{n-K}(y, z),$$

and since

$$\int_{\mathbf{R}^K} dy_0 \cdots dy_{K-1}\, f(y_0) \cdots f(y_{K-1}) h(y_0, y_1, \dots, y_{K-1}) = C_1 \int j(y)\, dy$$

is finite, the first inequality holds.

To prove the second inequality, we use Lemma 6.2, giving that $1/h(z)$ is a constant multiple of $g(\hat{z})/\pi(\hat{z})$. □

Our final ingredient is the following lemma.

LEMMA 6.4. *The law of large numbers and the central limit theorem hold for irreducible positive recurrent Harris chains started in their stationary distributions.*

PROOF. The approach of Athreya and Ney (1978) to the study of recurrent Harris chains on a state space S is to enlarge the state space by adding one point α that is hit by the chain with probability 1. [See Section 5.6 of Durrett (1995) for more details.] The law of large numbers and the central limit theorem can then be proved as in the discrete case by considering successive visits to α. See, for example, Exercises 5.5 and 5.6 in Chapter 5 of Durrett (1995). □

THEOREM 6.1. *If we let* $\mu = \int dy\, \bar{\pi}(y) h(y)\, y_{K-1}$, *let* $\varepsilon > 0$ *and let*

$$\Omega_{N,\varepsilon} = \left\{ (x_0, \dots, x_{N-1}) : \left| \frac{1}{N} \sum_{i=0}^{N-1} x_i - \mu \right| > \varepsilon \right\},$$

then $P_N(\Omega_{N,\varepsilon}) \to 0$ *as* $N \to \infty$.

Here and in the next result, the finiteness of the moments μ and σ^2 follow from Lemma 6.2 and our assumption that $\int e^{\theta x} f(x)\, dx < \infty$ for $\theta \in (-\delta, \delta)$. In these two results, it would be enough to assume that f has finite mean and variance.

PROOF OF THEOREM 6.1. Results from the theory of Harris chains imply that $Q_N(\Omega_{N,\varepsilon}) \to 0$ as $N \to \infty$. Let $\delta > 0$ and pick c so that $\int_{-\infty}^{c} f(y)\, dy = \delta$. Repeating the proof of the K-dimensional analogue of (6.3) with c in place of b in the definitions of A_I, Y_i^I, Z_i shows that

$$\int_{A_I} \Lambda^n q^n(x, y)\, dy \le \left(\frac{\delta}{1 - \delta} \right)^{K - |I|} \int_{[c,\infty)^K} \Lambda^n q^n(x, y)\, dy,$$

where $|I|$ is the number of elements in I. Now take $\delta \le 1/2$ so that $\delta/(1-\delta) \le 2\delta$ and let $B_{N,\delta} = \{x_j \le c \text{ for some } N-k \le j < N\}$. It follows from the previous inequality, (2.5) and (2.3) that

$$P_N(B_{N,\delta}) \le K\delta \frac{P(G'_{N-K})}{P(G_N)} \le C_K\delta.$$

Let $A_\delta = \{x_j \le c \text{ for some } 0 \le j < K\}$. Translation invariance of P_N implies that $P_N(A_\delta) = P_N(B_{N,\delta})$. Lemma 6.3 and relation (6.12) imply that on $A_\delta^c \cap B_{N,\delta}^c$ we have $dP_N/dQ_N \le C_\delta$, so

$$P_N(\Omega_{N,\varepsilon}) \le P_N(A_\delta) + P_N(B_{N,\delta}) + C_\delta Q_N(\Omega_{N,\varepsilon}).$$

From this, it follows that

$$\limsup_{N\to\infty} P_N(\Omega_{N,\varepsilon}) \le 2C_K\delta.$$

Since δ is arbitrary, the proof is complete. □

By working harder with these ideas, we can get a central limit theorem. Let

$$S_N = \frac{1}{\sigma\sqrt{N}}\left(\sum_{i=0}^{N-1} x_i - N\mu\right).$$

THEOREM 6.2. *Let $\Omega_{N,s} = \{(x_0, \dots, x_{N-1}) : S_N \le s\}$. There is a constant σ^2 so that*

$$P_N(\Omega_{N,s}) \to P(\chi \le s),$$

where χ has the standard normal distribution.

PROOF. Our first step is to truncate. From the proof of Theorem 6.1, we see that

$$|P_N(\Omega_{N,s}) - P_N(\Omega_{N,s} \cap A_\delta^c \cap B_{N,\delta}^c)| \le 2C_K\delta. \tag{6.13}$$

Results from the theory of Harris chains imply that $Q_N(\Omega_{N,s}) \to P(\chi \le s)$. Our goal, then, is to transfer these results to P_N. Let

$$T_N = \sum_{k=0}^{N-1} X_k \quad \text{and} \quad T'_N = \sum_{k=N^{1/4}}^{N-N^{1/4}} X_k.$$

Stationarity of the X_k under Q_N implies that

$$Q_N(|T_N - T'_N|) \le 2N^{1/4} Q_N(|X_k|),$$

where we have used $Q_N(f)$ as shorthand for $\int f\, dQ_N$. Thus, if we let $S'_N = (T'_N - N\mu)/\sigma\sqrt{N}$, we have

$$Q_N(|S_N - S'_N| > \varepsilon) \to 0 \tag{6.14}$$

for all $\varepsilon > 0$. Let

$$U = (X_0, \dots, X_{K-1}), \qquad V_N = (X_{N-K}, \dots, X_{N-1}).$$

The convergence of the Markov chain $\bar{q}(x, y)$ to its stationary distribution implies that, as $N \to \infty$, $X_{N^{1/4}}$ is asymptotically independent of U. The Markov property then implies that U and S'_N are asymptotically independent. That is,

$$\big|Q_N(U \le u, S'_N \le s) - Q_N(U \le u)Q_N(S'_N \le s)\big| \to 0 \tag{6.15}$$

as $N \to \infty$. Let

$$\hat{q}(x, y) = \frac{\bar{\pi}(y)\bar{q}(y, x)}{\bar{\pi}(x)}$$

be the transition probability for the time-reversed chain. Since $\hat{q}(x, y)$ is an irreducible Harris chain and has stationary distribution $\bar{\pi}(x)$, it is positive recurrent. Repeating the previous argument for this chain shows that V_N and $X_{N-N^{1/4}}$ are asymptotically independent. From this, it follows that V_N and (U, S'_N) are asymptotically independent. Combining this observation with (6.15), we have

$$\begin{aligned} &\big|Q_N(U \le u, S'_N \le s, V_N \le v) \\ &\qquad - Q_N(U \le u)Q_N(S'_N \le s)Q_N(V_N \le v)\big| \to 0. \end{aligned}$$

Note that, dQ_N-almost surely,

$$\frac{dP_N}{dQ_N} = f_N(U, V_N)$$

as indicated in (6.12). Due to asymptotic independence,

$$\begin{aligned} &\Big|Q_N\Big(f_N(U, V_N)\mathbb{1}_{(A_\delta^c \cap B_{N,\delta}^c)}\mathbb{1}_{(S'_N \le s)}\Big) \\ &\qquad - Q_N\Big(f_N(U, V_N)\mathbb{1}_{(A_\delta^c \cap B_{N,\delta}^c)}\Big)Q_N(S'_N \le s)\Big| \to 0. \end{aligned} \tag{6.16}$$

Due to (6.14) and the boundedness on $f_N(U, V_N)$ on $A_\delta^c \cap B_{N,\delta}^c$,

$$\begin{aligned} &\Big|Q_N\Big(f_N(U, V_N)\mathbb{1}_{(A_\delta^c \cap B_{N,\delta}^c)}\mathbb{1}_{(S'_N \le s)}\Big) \\ &\qquad - Q_N\Big(f_N(U, V_N)\mathbb{1}_{(A_\delta^c \cap B_{N,\delta}^c)}\mathbb{1}_{(S_N \le s)}\Big)\Big|. \end{aligned} \tag{6.17}$$

From (6.16) and (6.17), it follows that

$$\big|P_N(A_\delta^c \cap B_{N,\delta}^c \cap \Omega_{N,s}) - P_N(A_\delta^c \cap B_{N,\delta}^c)Q_N(\Omega_{N,s})\big| \to 0.$$

Combining this with (6.13) and the fact that $P_N(A_\delta^c \cap B_{N,\delta}^c) \ge 1 - 2C_K\delta$, the desired result follows. $\square$

7. Maxima. In this section, we will prove results about the number of local maxima and the height of the global maximum. The key to this is the observation that there are "cut points" where all local maxima must have specified bits and this breaks the overall maximization problem into a large number of independent maximization subproblems. We begin with the case $K = 1$. To define a cut point in this case, we note that if

$$(\star) \quad \phi_{i-1}(a,1) + \phi_i(1,b) > \phi_{i-1}(a,0) + \phi_i(0,b) \qquad \text{for all } a,b \in \{0,1\},$$

then any local maximum must have a 1 in the ith position. If $(\star)$ holds, then we say that i is a cut point.

To compute the probability of this event, let

$$U_0 = \phi_{i-1}(0,1), \qquad U_1 = \phi_{i-1}(1,1), \qquad U_2 = \phi_i(1,0), \qquad U_3 = \phi_i(1,1),$$
$$V_0 = \phi_{i-1}(0,0), \qquad V_1 = \phi_{i-1}(1,0), \qquad V_2 = \phi_i(0,0), \qquad V_3 = \phi_i(0,1).$$

In terms of the new variables, the event in $(\star)$ can be expressed as

$$U_j + U_k > V_j + V_k \qquad \text{for all } (j,k) \in \{(0,2),(1,2),(0,3),(1,3)\}.$$

The events $E_{j,k} = \{U_j + U_k > V_j + V_k\}$ are increasing functions of independent random variables $U_0, U_1, U_2, U_3, -V_0, -V_1, -V_2, -V_3$, so that Harris's inequality [see, e.g., Kesten (1981), page 72] gives the following:

$$P\Bigg(\bigcap_{(j,k)} E_{j,k}\Bigg) \ge \prod_{(j,k)} P(E_{j,k}) = \tfrac{1}{16}.$$

To compute the exact probability, we note that $(\star)$ is equivalent to

$$\min\{U_2 - V_2, U_3 - V_3\} > \max\{V_0 - U_0, V_1 - U_1\}.$$

The four differences in the last equation are independent and identically distributed. There are $4! = 24$ possible relative orders for these random variables, exactly four of which give the desired equality, so $(\star)$ has probability $1/6$.

The concept of a cut point generalizes easily to $K > 1$. For example, when $K = 2$ we want

$$\begin{aligned}&\phi_{i-2}(a,b,1) + \phi_{i-1}(b,1,1) + \phi_i(1,1,c)\\ &\qquad > \phi_{i-2}(a,b,u) + \phi_{i-1}(b,u,v) + \phi_i(u,v,c)\end{aligned}$$

and

$$\begin{aligned}&\phi_{i-1}(b,1,1) + \phi_i(1,1,c) + \phi_{i+1}(1,c,d)\\ &\qquad > \phi_{i-1}(b,u,v) + \phi_i(u,v,c) + \phi_{i+1}(v,c,d)\end{aligned}$$

for all $a,b,c,d \in \{0,1\}$ and $(u,v) \in \{0,1\}^2 - \{(1,1)\}$. These inequalities guarantee that value at any point can be improved by flipping both the ith and the $(i+1)$st bits to 1. There are 64 inequalities here. In some cases, for example, $\phi_{i-2}(a,b,1)$ and $\phi_{i+1}(1,c,d)$, the variables on the left-hand side can also appear on the right, but when this occurs they can be subtracted from each side. This allows us to use Harris's inequality as before to conclude the probability of a cut point is at least 2^{-64}. A more careful computation can reduce the probability of a cut point to 2^{-16} or less, but for our purposes that is not important. The existence of a positive density of cut points allows us to prove three results, the first of which is as follows.

THEOREM 7.1. *Let M_N be the number of local maxima. There are constants μ_M and σ_M^2 so that* $(\log M_N - \mu_M N)/\sqrt{N}$ *converges in distribution to a normal with mean* 0 *and variance σ_M^2.*

PROOF. For simplicity, we give the details only for $K=1$. As the reader will see, the proof generalizes in a straightforward way to $K>1$ but becomes more tedious to write down. To take the limit as $N\to\infty$, it is convenient to define an infinite family of independent random variables $\phi_i(\eta)$ for $i\ge 0$ and $\eta\in\{0,1\}^{K+1}$ and then use appropriate N-tuples of random variables from this family to construct the finite systems.

Let $L_1, L_2, \dots$ be the location of the cut points at sites $j\ge 1$ with $j \bmod 3 = 1$. We consider sites that are 3 units apart so that the corresponding events $\{j$ is a cut point$\}$ become independent and $w_i/3 = (L_{i+1}-L_i)/3$ has a geometric distribution. Let

$$v_i = \big(\phi_{L_i-1}(1,1), \phi_{L_i-1}(0,1), \phi_{L_i}(1,1), \phi_{L_i}(1,0)\big).$$

The sequence (v_i) is i.i.d. and independent of w_i, so (v_i, w_i, v_{i+1}) is a positive recurrent Harris chain.

Let m_i be the number of sequences $\eta\in\{0,1\}^{[L_i,L_{i+1}]}$ with $\eta(L_i)=\eta(L_{i+1})=1$ that are local maxima; that is, the value of $\sum_{j\in[L_i,L_{i+1})}\phi_j(\eta_j,\eta_{j+1})$ is not improved by changing any one of the coordinates $j\in(L_i,L_{i+1})$. If we condition on (v_i,w_i,v_{i+1}), $i\ge 1$, then the m_i are independent, so (v_i,w_i,m_i,v_{i+1}), $i\ge 1$, is a positive recurrent Harris chain.

Let $J(N)=\max\{i : L_{i+1}<N\}$ and let m_0 be the number of local maxima in the interval $[L_{J(N)+1}, L_1]$ that wraps around. Then $M_N = m_0\prod_{i=1}^{J(N)} m_i$. If we let $w_0 = L_1 - L_{J(N)+1} \bmod N$ be the width of the wrap-around interval, then it is easy to see that w_0 is bounded in distribution by three times a sum of two independent geometric random variables. Since $m_0\le 2^{w_0-1}$, we can ignore the contribution of $\log m_0$ to $\log M_N$ in proving a central limit theorem.

Since $\log_2 m_i \le w_i$ and $w_i/3$ has a geometric distribution, we have $E(\log m_i)^\rho<\infty$ for all $\rho<\infty$. The strong law of large numbers for positive recurrent Harris chains implies

$$\frac{1}{n}\sum_{i=1}^{n}\log m_i \to E\log m_i \qquad \text{a.s.},$$

where the right-hand side is the expected value of $\log m_i$ under the stationary distribution. The law of large numbers for renewal sequences implies $J(N)/N \to 1/Ew_i$. Combining the last two results, we have

$$\frac{1}{N}\sum_{i=1}^{J(N)} \log m_i \to \frac{E\log m_i}{Ew_i} \qquad \text{a.s.}$$

To derive the central limit theorem, we now note that

$$\sum_{i=1}^{J(N)} \log m_i = \sum_{i=1}^{N/Ew_i} \log m_i + \left(J(N) - \frac{N}{Ew_i}\right) E\log m_i + o(\sqrt{N}),$$

where $o(\sqrt{N})$ indicates a term that, when divided by $\sqrt{N}$, converges in distribution to 0. Similar reasoning, using $\sum_{i=1}^{J(N)} w_i \approx N$, shows

$$J(N) - \frac{N}{Ew_i} = \frac{N - \sum_{i=1}^{N/Ew_i} w_i}{Ew_i} + o(\sqrt{N}).$$

Combining the last two results, we have

$$\sum_{i=1}^{J(N)} \log m_i - N\frac{E\log m_i}{Ew_i} = \sum_{i=1}^{N/Ew_i} u_i + o(\sqrt{N}),$$

where $u_i = \log m_i - w_i(E\log m_i/Ew_i)$ has $Eu_i = 0$. The result now follows from the central limit theorem for positive recurrent Harris chains. □

THEOREM 7.2. *Let H_N^* be the height of the global maximum. There are constants μ_{H^*} and $\sigma_{H^*}^2$ so that $(H_N^* - \mu_{H^*}N)/\sqrt{N}$ converges in distribution to a normal with mean* 0 *and variance $\sigma_{H^*}^2$.*

PROOF. Again, we give the details only for the case $K = 1$. Let h_i be the height of the global maximum of $\sum_{j\in[L_i,L_{i+1})} \phi_j(\eta_j, \eta_{j+1})$ over all sequences $\eta \in \{0, 1\}^{[L_i,L_{i+1}]}$ with $\eta(L_i) = \eta(L_{i+1}) = 1$.

PROPOSITION 7.1. *$P(h_i > x) \le Ce^{-ax}$, where $a > 0$ and $C < \infty$ are some constants.*

Once Proposition 7.1 is established (see also the remark following it), we have $E|h_i|^\rho < \infty$ for all $\rho < \infty$. The rest of the proof is identical to that of Theorem 7.1. All we have to do is to replace $\log m_i$ by h_i. □

PROOF OF PROPOSITION 7.1. Let E_m be the event that m is a cut point and let $M_j = \max\{\phi_j(u, v) : (u, v) \in \{0, 1\}\}$. From the definition, it is easy to see that the event E_m is independent of M_j with $j < m - 1$ and $j > m$. Suppose, without loss of generality, that $L_i = 0$ and break things down according to the value of $L_i = 3k$. In this case,

$$h_i \le S_0 + S_1 + S_2 \qquad \text{where } S_l = \sum_{j=0}^{k-1} M_{3j+l}.$$

In order for $h_i > x$, we must have some $S_l > x/3$. Thus, to prove our result, it is enough to show that there is some $\gamma > 0$ so that $\max_l E\exp(\gamma S_l) < \infty$.

If we condition on the event $F_{3k} = E_0 \cap E_3^c \cap \cdots \cap E_{3k-3}^c \cap E_{3k}$, then the random variables

$$M_0, M_1, (M_2, M_3), M_4, (M_5, M_6), \ldots, M_{3k-5}, (M_{3k-4}, M_{3k-3}), M_{3k-2}, M_{3k-1}$$

are independent. The distribution of $M_1, M_4, \ldots, M_{3k-2}$ is not affected by the conditioning. Conditional on F_{3k}, the pairs (M_{3j-1}, M_{3j}) have the same distribution as $((M_{m-1}, M_m)|E_m^c)$ while M_0 and M_{3k-1} have the same distribution as $(M_m|E_m)$ and $(M_{m-1}|E_m)$.

We have supposed that $\int e^{\theta x} f(x)\,dx < \infty$ for $\theta \in (-\delta, \delta)$. Let $p = P(E_m)$ and pick $\varepsilon > 0$ small enough so that $Ee^{\gamma M_j} < 1/(1-p)^{1/2}$ for all $\gamma \in [0, \varepsilon]$. Breaking things down according to the value of k,

$$E\bigl(\exp(\gamma S_1)\bigr) = \sum_{k=1}^{\infty} p(1-p)^k (Ee^{\gamma M_j})^k \le \frac{p(1-p)^{1/2}}{1-(1-p)^{1/2}}. \tag{7.1}$$

To bound S_0 and S_2, we note that if X is a random variable and G is an event with positive probability, then

$$P(X > x|G) \le \frac{P(X > x)}{P(G)} \wedge 1.$$

From this, it follows that if we choose $0 < \gamma \le \varepsilon$ sufficiently small, then

$$Ee^{\gamma(M_m|E_m)},\ Ee^{\gamma(M_{m-1}|E_m)},\ Ee^{\gamma(M_m|E_m^c)},\ Ee^{\gamma(M_{m-1}|E_m^c)} \le 1/(1-p)^{1/2}.$$

Summing as in (7.1), we have, for $l = 0, 2$, $E(\exp(\gamma S_l)) \le p(1-p)^{1/2}/(1-(1-p)^{1/2})$ and the proof is complete. □

REMARK. Let g_i be the height of the global minimum of $\sum_{j\in[L_i, L_{i+1})} \phi_j(\eta_j, \eta_{j+1})$ over all sequences $\eta \in \{0,1\}^{[L_i, L_{i+1}]}$ with $\eta(L_i) = \eta(L_{i+1}) = 1$. The proof of Proposition 7.1 easily extends to showing $P(g_i < -x) \le Ce^{-ax}$.

As another consequence of cut-point decomposition, we can show that the heights of local maxima on one landscape follow a normal distribution. Before formulating the result, we introduce additional notation. Let $(w_1, X^1), (w_2, X^2), \ldots$ be independent and identically distributed taking values in $\mathbf{Z}^+ \times \bigcup_{m=1}^{\infty} [0,1]^m$. Here, X^i is a list of the heights of the local maxima in $[L_i, L_{i+1}]$ and

$w_i = L_{i+1} - L_i$, $i \geq 1$. Let m_i be the dimension of the vector X^i. Note that, given $m_i = m$ and $w_i = w$, $(X_1^i, X_2^i, \dots, X_m^i)$ form a pairwise exchangeable sequence with no ties. That is, given w_i, m_i, for any two i_1, i_2, $1 \leq i_1 < i_2 \leq m_i$, the conditional distribution of $(X_{i_1}^i, X_{i_2}^1)$ is symmetric in its coordinates and has no mass on the diagonal. Let

$$a_i = \frac{1}{m_i} \sum_{j=1}^{m_i} X_j^i$$

be the average value of the coordinates of X^i. Due to pairwise exchangeability, $Ea_i = EX_1^i$. Let $J(N) = \max\{i : w_1 + \cdots + w_i < N\}$ and let ν_N be the distribution that assigns mass $1/m_1 \cdots m_{J(N)}$ to each of the sums

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{J(N)} (X_{\eta(i)}^i - a_i) \qquad \text{where } \eta(i) \in \{1, \dots, m_i\}.$$

Equivalently, for each i, let $\eta(i)$ be a uniform random variable on $\{1, \dots, m_i\}$, given m_i, and let ν_E be the distribution of $\sum_{i=1}^{J(N)} (X_{\eta(i)}^i - a_i)/\sqrt{N}$.

In either formulation, $\nu_N(\omega, A)$ depends on the realization of the variables ω that are used to construct the sequence of landscapes and in the second variable is a measure on $\mathbf{R}$.

THEOREM 7.3. *For almost every ω, as $N \to \infty$, $\nu_N(\omega, \cdot)$ converges weakly to a normal distribution with variance σ_E^2.*

PROOF. Since, for each $j \in \{1, \dots, m_i\}$, $X_j^i \in [g_i, h_i]$, we have the existence of all moments and

$$\frac{1}{(\log N)^2} \max_{1 \leq i \leq J(N)} \max_{1 \leq j \leq m_i} |X_j^i| \to 0 \qquad \text{as } N \to \infty. \tag{7.2}$$

Let G_j^N be the distribution that assigns mass $1/m_j$ to each point $(X_j^i - a_i)/\sqrt{N}$. Then G_j^N has mean 0 and variance v_j^N. The law of large numbers implies that, as $N \to \infty$, $\sum_{j=1}^{J(N)} v_j^N \to \sigma_E^2 > 0$. Using (7.2), we now see that the triangular array

$$\sum_{i=1}^{J(N)} \frac{X_{\eta(i)}^i - a_i}{\sqrt{N}} \tag{7.3}$$

satisfies the hypotheses of the Lindberg–Feller central limit theorem and the desired result follows.

To relate the variance of the limiting normal here to that in Theorem 6.2, note that if we replace a_i by Ea_i in (7.3), then the new quantity

$$\sum_{i=1}^{J(N)} \frac{X_{\eta(i)}^i - Ea_i}{\sqrt{N}} \tag{7.4}$$

is very close to the CLT scaled fitness S_N, which was studied in Theorem 6.2. These two quantities differ since (7.4) does not depend on $\Phi_i(\eta_i, \dots, \eta_{i+K})$ for $i > \sum_{j=1}^{J(N)} w_j$. It is easy to see that the difference between them is a term of order $o(1)$ as $N \to \infty$. Since $a_i \in [g_i, h_i]$, a_i has all moments. Note that

$$E(a_i X^i_{\eta(i)}) = E\left[a_i E\left(X^i_{\eta(i)} \middle| m_i, \sum_{j=1}^{m_i} X^i_j\right)\right] = E(a_i^2).$$

This implies that $\operatorname{cov}(a_i, a_i - X^i_{\eta(i)}) = 0$. Writing

$$\sum_{i=1}^{J(N)} \frac{X^i_{\eta(i)} - Ea_i}{\sqrt{N}} = \sum_{i=1}^{J(N)} \frac{X^i_{\eta(i)} - a_i}{\sqrt{N}} + \sqrt{\frac{J(N)}{N}} \sum_{i=1}^{J(N)} \frac{a_i - Ea_i}{\sqrt{J(N)}},$$

we have $\sigma_H^2 = \sigma_E^2 + EJ(1) \cdot \operatorname{var}(a_1)$. $\square$

REFERENCES

ATHREYA, K. and NEY, P. E. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.

DURRETT, R. (1995). *Probability: Theory and Examples*. Wadsworth & Brooks/Cole, Belmont, CA.

EVANS, S. N. and STEINSALTZ, D. (2002). Estimating some features of NK fitness landscapes. *Ann. Appl. Probab.* To appear.

FLYVBJERG, H. and LAUTRUP, B. (1992). Evolution in rugged fitness landscapes. *Phys. Rev. A* **46** 6714–6723.

KAUFFMAN, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford Univ. Press.

KAUFFMAN, S. A. and LEVIN, S. A. (1987). Towards a general theory of adaptive walks on rugged landscapes. *J. Theoret. Biol.* **128** 11–45.

KESTEN, H. (1981). *Percolation Theory for Mathematicians*. Birkhäuser, Boston.

MACKEN, C. A., HAGAN, P. S. and PERELSON, A. S. (1991). Evolutionary walks on rugged landscapes. *SIAM J. Appl. Math.* **51** 799–827.

MACKEN, C. A. and PERELSON, A. S. (1989). Protein evolution on rugged landscapes. *Proc. Natl. Acad. Sci. USA* **86** 6191–6195.

RIESZ, F. and SZ.-NAGY, B. (1990). *Functional Analysis*. Dover, New York.

TWEEDIE, R. (1974). R-theory for Markov chains on a general state space. I. Solidarity properties and R-recurrent chains. *Ann. Probab.* **2** 840–864.

WEINBERGER, E. D. (1989). A more rigorous derivation of some properties of uncorrelated fitness landscapes. *J. Theoret. Biol.* **134** 125–129.

WEINBERGER, E. D. (1991). Local properties of Kauffman's NK model: A tunably rugged energy landscape. *Phys. Rev. A* **44** 6399–6413.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
E-MAIL: limic@math.ubc.ca

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BC V6T 1Z2
CANADA
E-MAIL: limic@math.ubc.ca