Adv. Appl. Prob. **39,** 776–798 (2007)
Printed in Northern Ireland
© *Applied Probability Trust* 2007

ASYMPTOTICS FOR LOCAL MAXIMAL STACK SCORES WITH GENERAL LOOP PENALTY FUNCTION

NIELS RICHARD HANSEN,* *University of Copenhagen*

Abstract

A stack is a structural unit in an RNA structure that is formed by pairs of hydrogen bonded nucleotides. Paired nucleotides are scored according to their ability to hydrogen bond. We consider stack/hairpin-loop structures for a sequence of independent and identically distributed random variables with values in a finite alphabet, and we show how to obtain an asymptotic Poisson distribution of the number of stack/hairpin-loop structures with a score exceeding a high threshold, given that we count in a proper, declumped way. From this result we obtain an asymptotic Gumbel distribution of the maximal stack score. We also provide examples focusing on the computation of constants that enter in the asymptotic distributions. Finally, we discuss the close relation to existing results for local alignment.

Keywords: Extreme value theory; maximal free energy score; local stack; loop penalty; Poisson approximation; reflected random walk; RNA

2000 Mathematics Subject Classification: Primary 60G70
Secondary 60G50; 60F10

1. Introduction

In the attempt to understand the molecular structure of RNA molecules given the primary sequence of RNA nucleotides – a sequence from the alphabet $\{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{u}\}$ – a lot of work has been invested in the development of models and algorithms for correctly predicting the secondary structure of an entire RNA sequence [12], [16], [22]. The objective is to maximize a score function – typically minus the free energy – over the space of secondary structures. There has been less focus on the distribution of the optimal score for random RNA sequences. To this end, an interesting theoretical result can be found in [21]. Letting M_n denote the maximal score for a sequence of n independent and identically distributed (i.i.d.) random variables from the RNA alphabet $\{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{u}\}$ Xiong and Waterman [21] found the proper scaling of M_n to obtain strong limit results. More precisely, they showed that for a specific scoring mechanism there is a phase transition in the parameter space between a logarithmic growth phase and linear growth phase of M_n. It was, furthermore, conjectured in [21] that for parameters in the logarithmic phase the normalized, with maximal score $\theta^* M_n - \log(K^* n)$ for some θ^*, $K^* > 0$ asymptotically follows a Gumbel distribution. The conjecture is based upon an analogy to local alignment of two independent sequences of i.i.d. random variables. Moreover, Xiong and Waterman reported that a simulation seems to confirm the conjecture. The almost sure limit, $\lim_{n\to\infty} M_n / \log n$, necessarily equals $1/\theta^*$, and this limit can in turn be related to the log-moment generating functions for M_n, $n \geq 1$. The constant K^* is, however, not given

Received 27 January 2005; revision received 14 May 2007.

* Postal address: Department of Mathematical Sciences, Universitetsparken 5, DK-2100, Copenhagen Ø, Denmark.
Email address: richard@math.ku.dk

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

any representation and there is no theoretical results that confirm the conjecture. For local alignment the analogous – and likewise conjectured – asymptotic result plays an important role in the assessment of statistical significance of local alignments as implemented for instance in BLAST [1], [2]. A completely satisfactory, theoretical confirmation of this practice is, however, still lacking – except for gapless local alignment, see [8], or gapped local alignment with some control on the number of gaps, see, e.g. [19].

We offer a solution to a more modest problem for RNA structure than the conjecture in [21]. We will restrict our attention to the maximal score over stacks with a single hairpin loop, i.e. no internal loops, bulges, or multibranch loops are allowed in the structure; see [11] for definitions and a thorough treatment of secondary structure components. This greatly reduces the complexity of the problem and allows us, as for local, gapless alignment, to employ results from the theory of random walks. What we show is that by counting the number of stack/hairpin-loop structures with a high score in a suitably declumped way, we can obtain a Poisson limit by using the results in [4]. The major result is Theorem 1, which confirms the conjecture in [21] in our more restrictive setup. In one respect we are, however, capable of being more general than in [21], and that is in the choice of penalty function on the length of the hairpin loop. Where Xiong and Waterman [21] considered a linear penalty function, we handle a completely general penalty function. First of all we obtain a condition in terms of the penalty function for the theorem to hold. Second we also obtain rather explicit representations of the constant K^*, and we investigate through several examples how the choice of penalty function influences this constant.

A complementary approach that relies on Poisson approximation techniques similar to those used in the present paper was given by Reinert and Schbath [17]. Their results focused on the occurrence of words and word collections in stationary Markov chains. As an application of their general results, they investigated the occurrence of certain word collections that can form stack/hairpin-loop structures. In the present paper the set of stack/hairpin-loop structures whose score exceeds a threshold can also be understood simply as a collection of words. What we focus on, however, is an asymptotic scenario where we understand precisely how the probability of finding a word from the collection decays with the threshold. Moreover, the general phenomena that some words are self-overlapping, which forces Reinert and Schbath [17] to consider a compound Poisson approximation in general, is shown to have vanishing probability asymptotically in the setup of the present paper; see also Remark 2.

If we choose *not* to penalize the length of the hairpin loop, the result in Theorem 1 is no longer valid. We discuss in Section 7 this particular situation and its intimate connection to results for local, gapless alignment.

Two appendices are included. Appendix A contains some technical inequalities that can be formulated in a more general framework than explicitly needed in the paper. Appendix B summarizes the results from [10], upon which the present paper relies heavily.

2. Local stacks and stack scores

Let $(X_k)_{k\geq 1}$ be a sequence of random variables taking values in a finite set E, and let $f : E \times E \to \mathbb{R}$ and $g : \mathbb{N}_0 \to (-\infty, 0]$ be given functions. We define, for i and j and $\delta \geq 0$ satisfying $1 \leq i - \delta$, $i \leq j + 1$, and $j + \delta \leq n$, the random variables

$$S^{\delta}_{i,j} = \sum_{k=1}^{\delta} f(X_{i-k}, X_{j+k}).$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

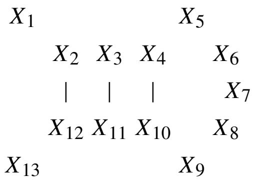

FIGURE 1: A graphical illustration of a stack/hairpin-loop structure for the random variables $X_1, \dots, X_{13}$. This structure corresponds to $(i, j, \delta) = (5, 9, 3) \in \mathcal{H}_n$ and the length of the hairpin-loop consisting of the variables $X_5, \dots, X_9$ is 5.

Then $S^{\delta}_{i,j} + g(j - i + 1)$ is the (loop-length penalized) score of the stack/hairpin-loop structure as given by (i, j, δ); see Figure 1. Let

$$\mathcal{H}_n = \{(i, j, \delta) \mid \delta \geq 0, 1 \leq i - \delta, i \leq j + 1, j + \delta \leq n\},$$

and define

$$\mathcal{M}_n = \max_{(i,j,\delta)\in\mathcal{H}_n} \{S^{\delta}_{i,j} + g(j - i + 1)\}$$

as the maximal, penalized score.

We define a matrix $(T_{i,j})_{i,j\geq 1}$ by $T_{i,j} = g(0)$ for $j < i$, $T_{i,i} = g(1)$, and recursively by

$$T_{i,j} = \max\{T_{i+1,j-1} + f(X_i, X_j), g(j - i + 1)\}$$

for $j > i$. Thus, we fill the matrix diagonally from the main diagonal towards the upper-right corner, and the final matrix becomes an upper triangular matrix. It follows that

$$T_{i,j} = \max_{\substack{(i',j',\delta)\in\mathcal{H}_n \\ i'-\delta=i, j'+\delta=j}} \{S^{\delta}_{i',j'} + g(j' - i' + 1)\}, \tag{1}$$

by verifying that the right-hand side fulfills the recursion.

If we introduce the set of upper triangular indices

$$\mathcal{H}^0_n = \{(i, j) \mid 1 \leq i, i \leq j + 1, j \leq n\},$$

it can be partitioned into diagonals in the following way. We say that (i, j) and (i', j') are on the same diagonal if $i' + j' = i + j$ (or equivalently $i' - i = j - j'$), and we call this partition I. Formally, we introduce an equivalence relation '$\sim$' on the set $\mathcal{H}^0_n$ by $(i, j) \sim (i', j')$ if $i' + j' = i + j$ and we can write $I = \mathcal{H}^0_n / \sim$. Since $i + j$ is constant for $(i, j) \in d$, $d \in I$, we can define $|d| = i + j$ taking any $(i, j) \in d$. We let $I_0 \subseteq I$ denote the set of diagonals where $|d|$ is odd and $I_1 \subseteq I$ the set of diagonals where $|d|$ is even. We note that for any $(i, j, \delta) \in \mathcal{H}_n$ with $(i, j) \in d$ then $|d|$ is even if and only if the hairpin-loop length $j - i + 1$ is odd. Also, note that $|d|$ takes values in $2, 3, \dots, 2n$, that I_0 contains $n - 1$ diagonals, and that I_1 contains the remaining n diagonals.

Introducing

$$\mathcal{M}^d_n = \max_{(i,j)\in d} T_{i,j}$$

as the maximum of the $T_{i,j}$-values along the diagonal d we find, due to (1) and the fact that the set of diagonals I forms a partition of $\mathcal{H}^0_n$, that

$$\mathcal{M}_n = \max_{d\in I} \mathcal{M}^d_n = \max_{1\leq i,j\leq n} T_{i,j}.$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

Thus, the maximum $\mathcal{M}_n$ can be computed as the maximum over the entries in the $T_{i,j}$-matrix. Defining, for $t \geq 0$, the counting variable

$$C_n(t) = \sum_{d \in I} \mathbf{1}_{\{\mathcal{M}_n^d > t\}},$$

(where $\mathbf{1}_{\{\cdot\}}$ is the indicator function) we obtain

$$(\mathcal{M}_n \leq t) = (C_n(t) = 0).$$

3. Results

We will assume that $(X_k)_{k\geq 1}$ is embedded in a doubly infinite sequence $(X_k)_{k\in\mathbb{Z}}$ of i.i.d. variables. We use the doubly infinite framework solely for notational convenience, for instance when introducing certain processes below. We let π denote the distribution of X_1 and assume (without loss of generality) that $\pi(x) > 0$ for all $x \in E$. We will assume that

$$f(x, y) > 0 \tag{2}$$

for some $x, y \in E$, and that f is *not* of the form

$$f(x, y) = f_1(x) + f_2(y)$$

for some $f_1, f_2 : E \to \mathbb{R}$. For convenience, we will also assume that f does not take values on a lattice; see Remark 1, below. That is, the set $\{f(x, y) \mid x, y \in E\}$ is *not* contained in a set of the form $\delta\mathbb{Z}$ for some $\delta > 0$.

We let

$$\mu = \sum_{x,y\in E} f(x, y)\pi(x)\pi(y)$$

denote the expectation of $f(X_{-1}, X_1)$ and we let

$$\varphi(\theta) = \sum_{x,y\in E} \exp(\theta f(x, y))\pi(x)\pi(y)$$

denote the Laplace transform of the distribution of $f(X_{-1}, X_1)$. It is a convex C^∞-function and $\mu = \partial_\theta\varphi(0)$. If $\mu < 0$ then there is a positive solution, θ^*, to the equation $\varphi(\theta) = 1$ since $\varphi(\theta) \to \infty$ for $\theta \to \infty$ due to (2). It is unique due to convexity.

Following Appendix B we introduce two stochastic processes by the recursive definitions:

$$T_n^0 = \max\{T_{n-1}^0 + f(X_{-n}, X_n), g(2n)\}, \qquad T_0^0 = g(0),$$

and

$$T_n^1 = \max\{T_{n-1}^1 + f(X_{-n}, X_n), g(2n+1)\}, \qquad T_0^1 = g(1).$$

According to Appendix B both processes $(T_n^i)_{n\geq 0}$, $i = 0, 1$, are random walks reflected in a barrier given by evaluating g in either the even or the odd integers. We note that for $d \in I_i$, $i = 0, 1$, the diagonal $(T_{i,j})_{(i,j)\in d}$ in the $T_{i,j}$-matrix has the same distribution as (a finite part of) the process $(T_n^i)_{n\geq 0}$.

By Theorem 4 in Appendix B it follows that if

$$M^i := \sup_{n\geq 0} T_n^i < \infty \quad \text{almost surely (a.s.)} \tag{3}$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

for $i = 0, 1$, then there are constants K_0^* and K_1^* such that

$$\mathrm{P}(M^i > x) \sim K_i^* \exp(-\theta^* x)$$

as $x \to \infty$ and for $i = 0, 1$. Define

$$K^* = K_0^* + K_1^*. \tag{4}$$

We should note that for (3) to hold it is obviously necessary (but not sufficient) that $\mu < 0$. With $g \equiv 0$, it follows, due to (2), that $\mathrm{P}(M^i = \infty) = 1$ even when $\mu < 0$, and therefore (3) can be viewed as a condition on the penalty function g; see also Section 4 and Appendix B.

Theorem 1. *Assume that $\mu < 0$, that $\theta^* > 0$ solves $\varphi(\theta) = 1$, that condition (3) is fulfilled, and that K^* is defined by (4). For $x \in \mathbb{R}$, let*

$$t_n = \frac{\log K^* + \log n + x}{\theta^*}, \tag{5}$$

then with $\mathcal{D}(C_n(t_n))$ denoting the distribution of $C_n(t_n)$ and $\|\cdot\|$ denoting the total variation norm, it holds that

$$\|\mathcal{D}(C_n(t_n)) - \mathrm{Poi}(\exp(-x))\| \to 0$$

for $n \to \infty$. In particular,

$$\mathrm{P}(\mathcal{M}_n \le t_n) \to \exp(-\exp(-x))$$

for $n \to \infty$.

Remark 1. To give a precise asymptotic distribution for $\mathcal{M}_n$ if f takes lattice values, for example integer values, we need to assume that g also takes values on the same lattice. If f and g take integer values, say, and the greatest common divisor of $f(x, y)$ for $x, y \in E$ is 1, then Theorem 1 holds under the same assumptions but with the following modifications. With $x_n \in [0, \theta^*)$ being given as $x_n = \theta^*(t_n - \lfloor t_n \rfloor)$, where $\lfloor\cdot\rfloor$ denotes the integer part function, then

$$\|\mathcal{D}(C_n(t_n)) - \mathrm{Poi}(\exp(-x + x_n))\| \to 0$$

for $n \to \infty$. In particular,

$$\mathrm{P}(\mathcal{M}_n \le t_n) - \exp(-\exp(-x + x_n)) \to 0$$

for $n \to \infty$. The proof of this is identical to the proof given below of Theorem 1 except that it relies on Remark 2.4 in [10] instead of our Theorem 4.

4. The constant K^*

The value of the constant K^* depends in a complicated way upon π, f, and g. We discuss here the representation of K_i^* as given in Appendix B, and some additional formulas for computing the quantities that enter in this representation.

Where

$$S_n = \sum_{k=1}^{n} f(X_{-k}, X_k)$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

is the random walk as in Appendix B, we introduce

$$D^i := \sup_{n\geq 0}\{g(2n+i) - S_n\}$$

and

$$C_i = \mathrm{E}^*(\exp(\theta^* D^i))$$

for $i = 0, 1$. Here, E^* denotes expectation under the exponentially changed measure P^* as introduced in Appendix B. According to Theorem 3 in Appendix B, $C_i < \infty$ if and only if $\mathrm{P}(M^i < \infty) = 1$. Note that (29), below, provides a useful criterion for verifying that $C_i < \infty$. If $C_i < \infty$ it follows, from Theorem 4 in Appendix B, that

$$K_i^* = \mathrm{E}^*(\exp(\theta^* D^i))\,\mathrm{E}^*(\exp(-\theta^* B)) = C_i C, \tag{6}$$

where B is a positive random variable. From the definition of the distribution of B in Appendix B the second factor in (6), $C = \mathrm{E}^*(\exp(-\theta^* B))$, does not depend upon g but only upon π and f. We conclude that

$$K^* = (C_0 + C_1)C.$$

There are several ways to represent and compute C. With $S_n^+ = \max\{S_n, 0\}$, [18, Corollary 8.45] gives

$$\begin{aligned} C &= \frac{\exp\{-\sum_{n=1}^{\infty}(1/n)\,\mathrm{E}^*(\exp(-\theta^* S_n^+))\}}{\theta^*\pi^*(f)} \\ &= \frac{\exp\{-\sum_{n=1}^{\infty}(1/n)[\mathrm{E}(\exp(\theta^* S_n);\, S_n \leq 0)) + \mathrm{P}(S_n > 0)]\}}{\theta^*\pi^*(f)}, \end{aligned} \tag{7}$$

which is a consequence of the Spitzer–Baxter identities. Here, π^* is the distribution of (X_{-1}, X_1) under P^*. An integral representation can also be given; see [18, Theorem 8.51], and the subsequent remarks.

Equation (28) in Appendix B gives a series representation of C_i for $i = 0, 1$, which we will use in the examples below. The formula is not analytically tractable but it can be used in combination with simulations. For a linear penalty function g we can obtain another, analytically more tractable, formula for computing C_i. If $g(n) = \alpha n$ for $\alpha < 0$, then

$$\tilde{S}_{i,n} = g(2n+i) - S_n = 2\alpha n - S_n + \alpha i$$

is a random walk (starting in αi) and D^i is thus the maximum of a random walk. We may first note that for the linear penalty function, (29) implies that $C_i < \infty$ whenever $\alpha < 0$ and thus in turn that (3) holds. It follows, from the Spitzer–Baxter identity, see [7, Theorem VIII.3.2], that

$$C_i = \exp(\theta^*\alpha i)\exp\left\{\sum_{n=1}^{\infty}\frac{1}{n}[\mathrm{E}^*(\exp(\theta^*\tilde{S}_{0,n}^+)) - 1]\right\}. \tag{8}$$

5. Examples

We consider three examples in detail and focus on the value and computation of K^*. More precisely, we focus on the computation of C_i for $i = 0, 1$ since this is a novel problem. Computing $C = \mathrm{E}^*(\exp(-\theta^* B))$ is in general not straightforward either, but it is a more

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

classical problem; see [7] and [18]. Equation (7) will work for the purpose of the examples we will consider, although in general it can be hard to compute the terms in the sum. Moreover, the constant C is only related to the random walk $(S_n)_{n\geq 0}$ and does not depend upon the penalty function g, whereas the factors C_i, $i = 0, 1$, represent the effect of the penalty function.

Common to all three examples below is the score function f, which is taken as

$$f(x, y) = \begin{cases} \log \dfrac{p}{p_0} & \text{if } x = y, \\ \log \dfrac{1-p}{1-p_0} & \text{if } x \neq y, \end{cases}$$

where $p_0 = \sum_{x\in E} \pi(x)^2$ and $p_0 < p < 1$. We note that (if $0 < p_0 < 1$) then

$$\mu = p_0 \log \frac{p}{p_0} + (1 - p_0) \log \frac{1-p}{1-p_0} < 0,$$

and we also find that $\theta^* = 1$. Moreover, the simplicity will allow for some rather explicit expressions. In all the examples below we will also take $p_0 = \frac{1}{2}$. The function f is seen to be nonlattice if and only if $\log(p/p_0)$ and $\log((1-p)/(1-p_0))$ are linearly independent over $\mathbb{Q}$. In other words, f is nonlattice if and only if there are no integer solutions to the equation

$$n \log \frac{p}{p_0} + m \log \frac{1-p}{1-p_0} = 0.$$

This provides a usable – though potentially complicated – criterion for checking whether f is lattice or not. There does not seem to be a simpler way of determining which ps and p_0s give rise to a lattice f.

Let

$$F_{n,p}(k) = \sum_{m=0}^{k} \binom{n}{m} p^m (1-p)^{n-m}$$

denote the distribution function for the binomial distribution with parameters (n, p) and let $\overline{F}_{n,p}(k) = 1 - F_{n,p}(k)$. Then with

$$n(p, p_0) = \left\lfloor \frac{n \log((1-p_0)/(1-p))}{\log(p(1-p_0)/p_0(1-p))} \right\rfloor,$$

we find that

$$\mathrm{E}(\exp(S_n); S_n \leq 0) = F_{n,p}(n(p, p_0))$$

and

$$\mathrm{P}(S_n > 0) = \overline{F}_{n,p_0}(n(p, p_0)).$$

This gives

$$C = \frac{\exp\{-\sum_{n=1}^{\infty} (1/n)[F_{n,p}(n(p, p_0)) + \overline{F}_{n,p_0}(n(p, p_0))]\}}{p \log(p/p_0) + (1-p) \log((1-p)/(1-p_0))}. \tag{9}$$

Example 1. Consider the linear penalty function $g(n) = \alpha n$, $\alpha < 0$, for which we know that (3) holds. Letting

$$n'(\alpha, p, p_0) = \left\lfloor \frac{n(\log((1-p)/(1-p_0)) - \alpha)}{\log(p_0(1-p)/p(1-p_0))} \right\rfloor,$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

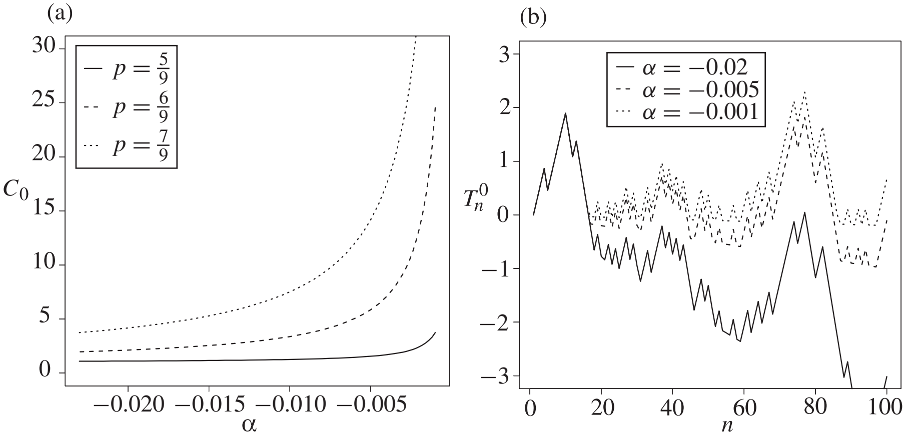

FIGURE 2: Considering the linear penalty function $g(n) = \alpha n$, $\alpha < 0$, we see (a) C_0 as a function of α for three different choices of p and $p_0 = \frac{1}{2}$. We also show examples of sample paths for the process $(T_n^0)_{n\geq 0}$ (b) for different choices of α, $p = \frac{2}{3}$, and $p_0 = \frac{1}{2}$.

TABLE 1: The value of C with $p_0 = \frac{1}{2}$, computed using (9), decreases as a function of the parameter p, but so does μ. Thus, increasing p results in a random walk $(S_n)_{n\geq 0}$ with a larger negative drift and a smaller value of C.

p	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
μ	−0.005	−0.020	−0.047	−0.087	−0.144	−0.223	−0.337	−0.511
C	0.96	0.92	0.89	0.86	0.83	0.80	0.78	0.76

we find, from (8), that

$$C_i = \exp(2\alpha i)\exp\left\{\sum_{n=1}^{\infty}\frac{1}{n}[\exp(2\alpha n)F_{n,p_0}(n'(\alpha, p, p_0)) + \overline{F}_{n,p}(n'(\alpha, p, p_0)) - 1]\right\}.$$

Using this formula, in Figure 2 we plot a graph of C_0 as a function of α for three different choices of p (and with $p_0 = \frac{1}{2}$). The infinite sum is truncated to 1000 terms. A small α results in larger values of M^0 in general and there is a corresponding increase in C_0. The effect of changing p is also seen. Larger values of p result in a process with larger fluctuations, which increases the effect of the penalty function on the value of M^0, and we see the corresponding increase in the value of C_0. Thus, the effect of increasing p goes in the opposite direction for C_0 compared C, which decreases for increasing p; see Table 1. The behavior of C_1 parallels C_0.

Example 2. An alternative to the linear penalty function is a piecewise linear penalty function, where

$$g(n) = \alpha \max\{n - n_0, 0\}$$

for $\alpha < 0$ and $n_0 \in \mathbb{N}_0$. Trivially, (29) implies that $C_i < \infty$ and thus in turn that (3) holds. Taking $n_0 = 0$ we obtain the linear penalty function, but for $n_0 > 0$ we obtain zero penalty up to n_0 and then the linear penalty from that point. We could choose such a penalty function if we want to favor small loops in a nonlinear way. Figure 3 shows, based on simulations

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

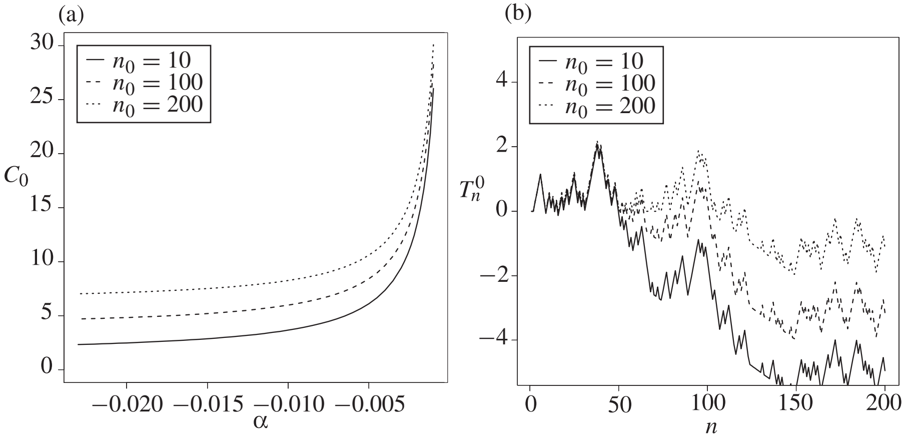

FIGURE 3: Considering the piecewise linear penalty function $g(n) = \alpha \max\{n - n_0, 0\}$, $\alpha < 0, n_0 \in \mathbb{N}_0$, we see (a) C_0 as a function of α for different choices of n_0 and with $p = \frac{2}{3}$ and $p_0 = \frac{1}{2}$. We also show examples of sample paths for the process $(T_n^0)_{n\geq 0}$ (b) for different choices of n_0.

of D_n^0 (under P) and the representation (28) of C_0, the value of C_0 as a function of α for $n_0 = 10, 100, 200$ and with $p = \frac{2}{3}$ and $p_0 = \frac{1}{2}$. The infinite sum is truncated to 10 000 terms, and since all terms in (28) are positive this gives the upper bound

$$\sum_{n=10\,001}^{\infty} \exp(g(n)) = \frac{\exp(\alpha(10\,001 - n_0))}{1 - \exp(\alpha)}$$

on the error due to the truncation. For $n_0 = 200$ and $\alpha = -0.001$ this upper bound is equal to 0.055. In addition to this (small) truncation error, there is the random error due to the simulations. We used 500 i.i.d. replications of D_n^0 to estimate C_0 using (28) and the largest estimate of the standard error (obtained for $n_0 = 200$ and $\alpha = -0.001$) was just below 0.25. The same conclusion about the effect of α as for the linear penalty function holds, i.e. small values of α give the largest value of C_0. The effect of n_0 is also clear. We see, as we would anticipate, that C_0 increases for increasing n_0.

Example 3. Consider the logarithmic penalty function $g(n) = \alpha \log n$, $\alpha < -1$. To show that (3) holds using (29) we need $\alpha < -1$. We can show that if $\alpha > -1$ then $\mathrm{P}(M^i = \infty) = 1$; see [10, Example 2.7]. Figure 4 shows three sample paths for $\alpha = -2, -1.1, -0.5$ (where $C_i = \infty$). We can hardly see the effect of the penalty function when $\alpha = -2$. Figure 4 also shows C_0, computed as in Example 2 using simulations and (28), as a function of α for various values of p and $p_0 = \frac{1}{2}$. For this computation we truncated the sum at 4000 terms, which for $\alpha = -1.5$ yields the upper bound 0.032 on the truncation error. However, for $\alpha = -1.1$ this upper bound is 4.36, and the improvement by increasing the number of terms to 10 000, say, is not serious – the upper bound is then 3.98. This is due to the slow convergence of the series

$$\sum_{n=1}^{\infty} \exp(\alpha \log n) = \sum_{n=1}^{\infty} n^{\alpha},$$

especially for α close to -1, which also renders the representation (28) less useful. Despite this deficiency, we see that the value of C_0 is close to 1 for most values of α, and only when p

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

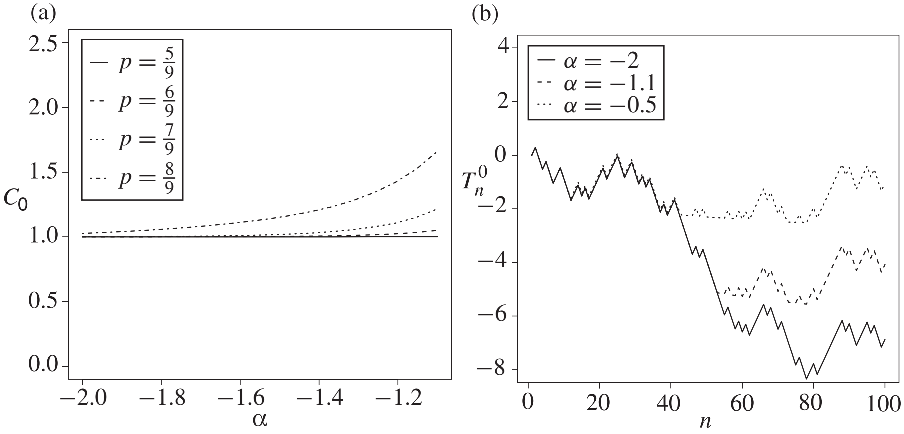

FIGURE 4: Considering a logarithmic penalty function $g(n) = \alpha \log(n)$ we need $\alpha < -1$ for C_0 to be finite, and we see (a) C_0 as a function of α for different choices of p and $p_0 = \frac{1}{2}$. We also see examples of sample paths for the process $(T_n^0)_{n\geq 0}$ (b) for different choices of α, $p = \frac{2}{3}$, and $p_0 = \frac{1}{2}$.

approaches 1 or α is close to the critical value -1 does C_0 take values that are notably larger than 1. In conclusion, it seems that the initial steepness of a logarithmic penalty function when $|\alpha|$ is sufficiently large almost completely prevents hairpin-loops in random sequences, which leads to $C_i \simeq 1$, $i = 0, 1$. When α approaches -1 (or exceeds -1) the hairpin-loops are not as heavily penalized, but then we might actually question the applicability of the asymptotic results, and for $\alpha > -1$ Theorem 1 can no longer be applied.

6. Proofs

The proof of Theorem 1 is an application of [3, Theorem 1], which in turn is a consequence of the Chen–Stein method. The sum of the indicator variables $\mathbf{1}_{\{M_n^d > t_n\}}$ is unfortunately not directly suitable for an application of [3, Theorem 1]. We will therefore band-limit the upper triangular matrix $(T_{i,j})_{i,j}$ before taking the maximum for each $d \in I$. For a given band-limiting sequence $b = (b_n)_{n\geq 0}$ we define, for $d \in I$,

$$\mathcal{M}_n^{d,b} = \max_{\substack{(i,j)\in d \\ |j-i|\leq 2b_n}} T_{i,j}$$

as the band-limited maximum of the diagonal given by d. Throughout we will assume that $2b_n \leq n$ and that

$$\lim_{n\to\infty} b_n^{-1} \log n = \lim_{n\to\infty} n^{-\varepsilon} b_n = 0 \tag{10}$$

for all $\varepsilon > 0$. The band-limitation is used in the present paper as a technical tool with the sole purpose of proving that $\sum_{d\in I} \mathbf{1}_{\{M_n^d > t_n\}}$ asymptotically follows a Poisson distribution. There is, however, an additional, practical gain, since we will show that the sum of the band-limited, diagonal maxima that exceeds t_n asymptotically is equal to $\sum_{d\in I} \mathbf{1}_{\{M_n^d > t_n\}}$. Thus, from a practical point of view, we really only need to compute the values of $T_{i,j}$ up to the band-limit, and this can be a serious, computational advantage.

We define, for $d \in I$,

$$V_d = \mathbf{1}_{\{\mathcal{M}_n^{d,b} > t_n\}}.$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

In the framework of [3, Theorem 1] we need to define a neighborhood of dependence for the variable V_d, i.e. a subset $B_d \subseteq I$ such that, for $d' \notin B_d$, V_d and $V_{d'}$ are independent. If we, for $d \in I$, define

$$B_d = \{d' \in I \mid \| |d| - |d'| \| \leq 4b_n\},$$

it is a simple matter to verify that due to the band limitation the variables V_d and $V_{d'}$ are indeed independent if $d' \notin B_d$.

With these definitions we rephrase [3, Theorem 1] in a suitable form.

Theorem 2. *If*

$$\lambda_n := \sum_{d \in I} \mathrm{E}(V_d) \to \lambda \tag{11}$$

for $n \to \infty$, and

$$\beta_{1,n} = \sum_{d \in I, d' \in B_d} \mathrm{E}(V_d)\,\mathrm{E}(V_{d'}) \to 0, \tag{12}$$

$$\beta_{2,n} = \sum_{d \in I, d' \in B_d, d \neq d'} \mathrm{E}(V_d V_{d'}) \to 0, \tag{13}$$

for $n \to \infty$, then

$$\left\| \mathcal{D}\left(\sum_{d \in I} V_d\right) - \mathrm{Poi}(\lambda) \right\| \to 0.$$

In fact, the bound

$$\left\| \mathcal{D}\left(\sum_{d \in I} V_d\right) - \mathrm{Poi}(\lambda_n) \right\| \leq 2(\beta_{1,n} + \beta_{2,n})$$

always holds.

We verify conditions (11), (12), and (13) in the following series of lemmas.

Lemma 1. *Under the assumptions given by (10), we have*

$$\lambda_n = \sum_{d \in I} \mathrm{E}(V_d) \to \exp(-x)$$

for $n \to \infty$.

Proof. The process $(T_n^i)_{n \geq 0}$ has the representation

$$T_n^i = S_n + \max_{0 \leq k \leq n} \{g(2k + i) - S_k\},$$

as discussed in Appendix B. Defining

$$\tau_i(u) = \inf\{n \geq 0 \mid T_n^i > u\},$$

we see that $\tau_i(u)$ is a stopping time with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$ defined in Appendix B, and

$$\begin{aligned} S_{\tau_i(u)} &= u + T_{\tau_i(u)} - u + S_{\tau_i(u)} - T_{\tau_i(u)} \\ &\geq u - \max_{0 \leq k \leq \tau_i(u)} \{g(2k + i) - S_k\} \\ &\geq u - \sup_{n \geq 0}\{g(2n + i) - S_n\} \\ &= u - D^i; \end{aligned}$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

hence, for $A \in \mathcal{F}_{\tau_i(u)}$ with $A \subseteq (\tau_i(u) < \infty)$, we obtain, from (26),

$$\mathrm{P}(A) \leq \exp(-\theta^* u)\,\mathrm{E}^*(\exp(\theta^* D_i);\, A). \tag{14}$$

Since $T_n^i - S_n = \max_{0 \leq k \leq n}\{g(2k+i) - S_k\}$ converges P^*-a.s. to the finite limit D^i, it also follows, from nonlinear renewal theory, that

$$\frac{\tau_i(u)}{u} \to \frac{1}{\mu^*}$$

in P^*-probability for $u \to \infty$; see [18, Chapter IX] and, in particular, [18, Lemma 9.13]. Consequently,

$$\frac{\tau(t_n)}{b_n} = \frac{\tau(t_n)}{\log n}\frac{\log n}{b_n} \to 0$$

in P^*-probability for $n \to \infty$, since $t_n \sim \log n/\theta^*$; also, since $\mathrm{E}^*(\exp(\theta^* D^i)) < \infty$ by assumption we obtain, in particular,

$$r_i(n) := \mathrm{E}^*(\exp(\theta^* D^i);\, \tau(t_n) > b_n) \to 0$$

for $n \to \infty$, which will prove to be useful.

Consider a diagonal, $d \in I_i$, with $2b_n \leq |d| \leq n - 2b_n$. Then we have

$$\mathrm{E}(V_d) = \mathrm{P}(\mathcal{M}_n^{d,b} > t_n) = \mathrm{P}(M_{b_n}^i > t_n),$$

where

$$M_{b_n}^i = \max_{0 \leq k \leq b_n} T_k^i.$$

We see that

$$\mathrm{P}(M^i > t_n) = \mathrm{P}(M_{b_n}^i > t_n) + \mathrm{P}(M_{b_n}^i \leq t_n,\, M^i > t_n)$$

and, since $(M_{b_n}^i \leq t_n,\, M^i > t_n) = (b_n < \tau_i(t_n) < \infty)$, it follows, from (14) (and the fact that $\mathrm{P}^*(\tau_i(u) < \infty) = 1$), that

$$\mathrm{P}(M_{b_n}^i \leq u,\, M^i > u) \leq \exp(-\theta^* t_n)\,\mathrm{E}^*(\exp(\theta^* D_i);\, \tau_i(t_n) > b_n) = K_1 n^{-1} r_i(n),$$

where $K_1 = \exp(-x)/K^*$. Using (14) again, we find, for any $d \in I_i$, that

$$\mathrm{E}(V_d) \leq \mathrm{P}(M^i > t_n) = \mathrm{P}(\tau_i(t_n) \leq \infty) \leq K_2^i n^{-1},$$

with $K_2^i = C_i \exp(-x)/K^*$. Since

$$(n-1)\,\mathrm{P}(M^0 > t_n) + n\,\mathrm{P}(M^1 > t_n) \sim (n-1)K_0^* \exp(-\theta^* t_n) + nK_1^* \exp(-\theta^* t_n) \to \exp(-x)$$

as $n \to \infty$, we see that

$$r_2(n) := |(n-1)\,\mathrm{P}(M^0 > t_n) + n\,\mathrm{P}(M^1 > t_n) - \exp(-x)| \to 0$$

as $n \to \infty$. Summing up, we have

$$\begin{aligned}\left|\sum_{d \in I} \mathrm{E}(V_d) - \exp(-x)\right| &\leq \sum_{d \in I_0} |\mathrm{E}(V_d) - \mathrm{P}(M^0 > t_n)| + \sum_{d \in I_1} |\mathrm{E}(V_d) - \mathrm{P}(M^1 > t_n)| \\ &\quad + |(n-1)\,\mathrm{P}(M^0 > t_n) + n\,\mathrm{P}(M^1 > t_n) - \exp(-x)| \\ &\leq K_1 r_0(n) + 4b_n K_2^0 n^{-1} + K_1 r_1(n) + 4b_n K_2^1 n^{-1} + r_2(n) \\ &\to 0\end{aligned}$$

as $n \to \infty$, by (10).

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

Lemma 2. *Under the assumptions given by (10), we have*

$$\beta_{2,n} = \sum_{d \in I, d' \in B_d} \mathrm{E}(V_d)\,\mathrm{E}(V_{d'}) \to 0$$

as $n \to \infty$.

Proof. Using the bound

$$\mathrm{E}(V_d) \le K_2^i n^{-1}$$

for $d \in I_i$, as found in the proof of Lemma 1, together with the fact that $|B_d| \le 8b_n$, we find that

$$\sum_{d \in I, d' \in B_d} \mathrm{E}(V_d)\,\mathrm{E}(V_{d'}) \le 8b_n n(K_2^0 + K_2^1)n^{-2} \to 0$$

as $n \to \infty$, by (10).

Lemma 3. *Under the assumptions given by (10), we have*

$$\beta_{3,n} = \sum_{d \in I, d' \in B_d, d \ne d'} \mathrm{E}(V_d V_{d'}) \to 0$$

as $n \to \infty$.

The proof of this lemma is subdivided into a couple of additional lemmas. First we will formulate and prove an exponential inequality, which is a rather standard consequence of the Azuma–Hoeffding inequality. To do so, we find it beneficial to introduce a few graph constructions.

Suppose that $(V_1, \mathcal{E}_1)$ and $(V_2, \mathcal{E}_2)$ are two graphs with finite vertex sets $V_1, V_2 \subseteq \mathbb{N}_0$. We will assume that each vertex has precisely one edge (and there are no loops), i.e. the graphs form *perfect matchings* of the vertex sets. We will in addition let $\mathcal{E}_\infty$ denote a set of edges that form a perfect matching on $\mathbb{N}_0$ such that $\mathcal{E}_1 \subseteq \mathcal{E}_\infty$. The existence of such an extension of the perfect matching on V_1 is clear, and the purpose is solely to make some formulations more convenient.

We let $V = V_1 \cap V_2$ denote the common vertex set and let $\boldsymbol{x} = (x_k)_{k \ge 0}$ denote an infinite sequence of elements from E. Introduce, for $k \in V$,

$$f_k^1(\boldsymbol{x}) = \begin{cases} f(x_k, x_m) & \text{if } \{k, m\} \in \mathcal{E}_1 \text{ and } k < m, \\ f(x_m, x_k) & \text{if } \{k, m\} \in \mathcal{E}_1 \text{ and } m < k. \end{cases}$$

Define f_k^2 for $k \in V$ likewise but based on the edge set $\mathcal{E}_2$ instead. We then define

$$s_V(\boldsymbol{x}) = \sum_{k \in V} f_k^2(\boldsymbol{x}) - f_k^1(\boldsymbol{x}).$$

In order to apply the Azuma–Hoeffding inequality, we need to verify that s_V has a certain Lipschitz property. If $\boldsymbol{x} = (x_k)_{k \ge 0}$ and $\boldsymbol{y} = (y_k)_{k \ge 0}$ satisfy that $x_k = y_k$ for all $k \notin \{k_1, m_1, k_2, m_2\}$, where $\{k_1, m_1\}, \{k_2, m_2\} \in \mathcal{E}_\infty$, then

$$|s_V(\boldsymbol{x}) - s_V(\boldsymbol{y})| \le 16 \max_{x, y \in E} |f(x, y)|. \tag{15}$$

Indeed, there are at most four terms in the two sums that can differ and each of these differences can trivially be bounded by $4 \max_{x,y \in E} |f(x, y)|$.

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

Lemma 4. *Let $\widetilde{\mathrm{P}}$ denote a probability measure such that under this measure $(X_k)_{k\geq 0}$ forms a sequence of random variables where (X_k, X_m) for $\{k, m\} \in \mathcal{E}_\infty$ (and $k < m$) are independent. Let $S_V = s_V((X_k)_{k\geq 0})$ and let $\xi_V = \widetilde{\mathrm{E}}(S_V)$ denote the expectation of S_V under $\widetilde{\mathrm{P}}$. If $\xi_V < 0$ it holds that*

$$\widetilde{\mathrm{P}}(S_V \geq 0) \leq \exp\biggl(-\frac{\xi_V^2}{2|V|\eta^2}\biggr),$$

with

$$\eta = 16 \max_{x,y\in E} |f(x, y)|,$$

and with $|V|$ denoting the number of elements in V.

Proof. By assumption, for each $k \in V$ there exist unique $m_1(k), m_2(k), m_3(k) \in \mathbb{N}_0$ such that $\{k, m_2(k)\} \in \mathcal{E}_2$ and $\{k, m_1(k)\}, \{m_2(k), m_3(k)\} \in \mathcal{E}_\infty$. We define a filtration $(\mathcal{F}_k)_{k\in V}$ by

$$\mathcal{F}_k = \sigma\{X_{k'}, X_{m_1(k')}, X_{m_2(k')}, X_{m_3(k')},\ k' \in V, k' \leq k\}$$

and, for $k \in V$, the random variable

$$Z_k = \mathrm{E}(S_V - \xi_V \mid \mathcal{F}_k).$$

Then $(Z_k, \mathcal{F}_k)_{k\in V}$ is a martingale with mean 0. We say that $k' \in V$ is a predecessor of $k \in V$ if k' is the largest element in V strictly smaller than k. We let -1 be the predecessor of the smallest element in V and $\mathcal{F}_{-1}$ the trivial σ-algebra so that $Z_{-1} = \mathrm{E}(S_V - \xi_V \mid \mathcal{F}_{-1}) = 0$. For all $k \in V$ with predecessor k', if we have

$$|Z_k - Z_{k'}| \leq c_k$$

for some constants c_k, then the Azuma–Hoeffding inequality [15, Lemma 5.1] reads, for $k \in V$ and all $\lambda > 0$,

$$\widetilde{\mathrm{P}}(Z_k \geq \lambda) \leq \exp\biggl(-\frac{\lambda^2}{2\sum_{m\in V, m\leq k} c_m^2}\biggr).$$

It is a direct consequence of the independence assumption and the Lipschitz property of s_V, as expressed by (15), that

$$|Z_k - Z_{k'}| = |\mathrm{E}(S_V \mid \mathcal{F}_k) - \mathrm{E}(S_V \mid \mathcal{F}_{k'})| \leq \eta = 16 \max_{x,y\in E} |f(x, y)|.$$

To complete the proof let $m \in V$ be the largest element in V then $Z_m = S_V - \xi_V$ and, if $\xi_V < 0$, we have

$$\widetilde{\mathrm{P}}(S_V \geq 0) = \widetilde{\mathrm{P}}(S_V - \xi_V \geq -\xi_V) = \widetilde{\mathrm{P}}(Z_m \geq -\xi_V) \leq \exp\biggl(-\frac{\xi_V^2}{2|V|\eta^2}\biggr).$$

Lemma 5. *There exists an $\varepsilon > 0$ such that, for all $(i, j, \delta), (i', j', \delta') \in \mathcal{H}_n$ with (i, j) and (i', j') not on the same diagonal and $t \geq 0$, we have*

$$\mathrm{P}(S_{(i,j)}^{\delta} > t, S_{(i',j')}^{\delta'} > t) \leq \exp(-\theta^*(1 + \varepsilon)t). \tag{16}$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

Proof. We define the graph $(V_1, \mathcal{E}_1)$ by

$$V_1 = \{i-\delta, \dots, i-1, j+1, \dots, j+\delta\}, \qquad \mathcal{E}_1 = \{\{i-k, j+k\} \mid 1 \le k \le \delta\},$$

and $(V_2, \mathcal{E}_2)$ is defined likewise using (i', j', δ'). Then we see that the intersection, V, of the vertex sets corresponds precisely to the set of variables X_i that enter in both of the sums $S^{\delta}_{i,j}$ and $S^{\delta'}_{i',j'}$. This implies that if we define

$$V_0 = \{k \mid i-k \in V \text{ or } j+k \in V\}, \qquad V_0' = \{k \mid i'-k \in V \text{ or } j'+k \in V\},$$

together with

$$S_1 = \sum_{k \in V_0} f(X_{i-k}, X_{j+k}), \qquad S_1' = \sum_{k \in V_0'} f(X_{i'-k}, X_{j'+k}),$$

then $S_2 := S^{\delta}_{i,j} - S_1$ and $S_2' := S^{\delta'}_{i',j'} - S_1'$ are independent. Since there are at most $|V|$ terms in S_1 and S_1' it follows that if $|V| \le t(4\|f\|_\infty)^{-1}$, say, then independence and an exponential change of measure gives

$$\begin{aligned}
\mathrm{P}(S^{\delta}_{(i,j)} > t, S^{\delta'}_{(i',j')} > t) &\le \mathrm{P}\biggl(S_2 > \frac{3}{4t}, S_2' > \frac{3}{4t}\biggr) \\
&= \mathrm{P}\biggl(S_2 > \frac{3}{4t}\biggr)\mathrm{P}\biggl(S_2' > \frac{3}{4t}\biggr) \\
&\le \exp\biggl(-\frac{3}{2\theta^* t}\biggr).
\end{aligned}$$

In particular, (16) holds with $\varepsilon = \frac{1}{2}$.

Suppose, instead, that $|V| \ge t(4\|f\|_\infty)^{-1}$. With $S_V = s_V((X_k)_{k \ge 0})$ as above, we have

$$\mathrm{P}(S^{\delta}_{(i,j)} > t, S^{\delta'}_{(i',j')} > t) \le \mathrm{P}(S^{\delta}_{(i,j)} > t, S_V \ge 0) + \mathrm{P}(S^{\delta'}_{(i',j')} > t, S_V \le 0). \tag{17}$$

We introduce the probability measure $\mathrm{P}^*_{(i,j,\delta)}$ by

$$\frac{\mathrm{d}\mathrm{P}^*_{(i,j,\delta)}}{\mathrm{d}\mathrm{P}} = \exp(\theta^* S^{\delta}_{i,j}),$$

and then, considering the first term in (17), we find that

$$\begin{aligned}
\mathrm{P}(S^{\delta}_{(i,j)} > t, S_V \ge 0) &= \mathrm{E}^*_{(i,j,\delta)}(\exp(-\theta^* S^{\delta}_{i,j}); S^{\delta}_{(i,j)} > t, S_V \ge 0) \\
&\le \exp(-\theta^* t)\, \mathrm{P}^*_{(i,j,\delta)}(S_V \ge 0),
\end{aligned}$$

where $\mathrm{E}^*_{(i,j,\delta)}$ denotes expectation under $\mathrm{P}^*_{(i,j,\delta)}$. It follows that under $\mathrm{P}^*_{(i,j,\delta)}$ the distribution of the sequence of variables $(X_i)_{i \ge 0}$ is as follows:

- the variables (X_k, X_m) for $\{k, m\} \in \mathcal{E}_1$ with $k < m$ are i.i.d. with distribution π^*,
- the variables X_k for $k \notin V_1$ are i.i.d. with distribution π,
- $(X_k)_{k \in V_1}$ and $(X_k)_{k \notin V_1}$ are independent.

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

The probability measure π^* on $E \times E$, with marginals denoted by π_1^* and π_2^*, is defined in Appendix B. We draw two conclusions. First, for any extension of $\mathcal{E}_1$ to a perfect matching $\mathcal{E}_\infty$ of $\mathbb{N}_0$, it holds that (X_k, X_m) for $\{k, m\} \in \mathcal{E}_\infty$ with $k < m$ are independent. Second, since (i, j) and (i', j') are not on the same diagonal the expectation of each term in S_V can be bounded above by $\zeta := \max\{\pi_1^* \otimes \pi_2(f), \pi_1^* \otimes \pi_2^*(f)\} - \pi^*(f)$. Hence,

$$\mathrm{E}(S_V) = \xi_V \le \zeta|V|,$$

where $\zeta < 0$ according to Lemma 6 in Appendix A. The assumptions of Lemma 4 are fulfilled and we find that, since $|V| \ge t(4\|f\|_\infty)^{-1}$,

$$\mathrm{P}^*_{(i,j,\delta)}(S_V \ge 0) \le \exp\biggl(-\frac{\zeta^2|V|}{2\eta^2}\biggr) \le \exp\biggl(-\frac{\zeta^2 t}{8\|f\|_\infty\eta^2}\biggr) = \exp(-\theta^*\varepsilon t),$$

where $\varepsilon = \zeta^2(8\theta^*\|f\|_\infty\eta^2)^{-1}$. By a similar argument, we can deal with the other term in (17), and this completes the proof.

Proof of Lemma 3. For any $d \in I$ there are at most b_n^2 elements $(i, j, \delta) \in \mathcal{H}_n$ fulfilling that $(i, j) \in d$ and $j - i + 2\delta \le 2b_n$. Since V_d indicates that for one such (i, j, δ) we have $S_{i,j}^\delta > t_n$, it follows, from Lemma 5, that, for $d, d' \in I$ with $d \ne d'$, we have

$$\mathrm{E}(V_d V_{d'}) \le b_n^4 \exp(-\theta^*(1+\varepsilon)t_n).$$

Then, since $|I| = 2n - 1$, $|B_d| \le 8b_n$, and t_n is given by (5), we obtain

$$\begin{aligned}\sum_{d\in I, d'\in B_d, d\ne d'} \mathrm{E}(V_d V_{d'}) &\le 16 b_n n b_n^4 \exp(-\theta^*(1+\varepsilon)t_n)\\ &\le \tilde{K} b_n^5 n^{-\varepsilon}\\ &\to 0\end{aligned}$$

as $n \to \infty$, due to (10).

Remark 2. Lemma 3 essentially shows, given the assumptions of this paper, that, for a sequence of letters that reach a high score, the probability that a shift of the sequence reach an equally high score is of an asymptotically smaller order. The proof relies among other things on the assumption that f is *not* of the form

$$f(x, y) = f_1(x) + f_2(y). \tag{18}$$

In Lemma 5 with reference to Lemma 6 in Appendix A the assumption is used to establish that $\zeta < 0$. From the expression for ζ in the proof of Lemma 5 we see how ζ quantifies 'nonadditivity' of f and consequently determines the size of the ε we can take in Lemma 5. If f is of the, biologically quite nonsensical, form (18) we have $\zeta = 0$, and for such a function the proof breaks down. This is because parallel diagonals in the score matrix can then reach scores of asymptotically the same order, which forces us to use a declumping technique to obtain a Poisson limit result. This phenomena is closely related to the potential self-overlap of words as discussed in [17].

Proof of Theorem 1. Note that

$$\sum_{d\in I} V_d \le C(t_n).$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

It is then possible to show that $\mathrm{E}(C(t_n)) \to \exp(-x)$ by the same arguments as in the proof of Lemma 1, but it is actually sufficient to verify the easier result that $\limsup_{n\to\infty} \mathrm{E}(C(t_n)) \le \exp(-x)$. Indeed,

$$\begin{aligned}\mathrm{E}(C_n(t_n)) &= \sum_{d\in I} \mathrm{P}(\mathcal{M}_n^d > t_n)\\ &\le (n-1)\,\mathrm{P}(M^0 > t_n) + n\,\mathrm{P}(M^1 > t_n)\\ &\to \exp(-x)\end{aligned}$$

for $n \to \infty$. Then, by the coupling inequality and the fact that the random variables are integer valued, we have

$$\begin{aligned}\limsup_{n\to\infty}\bigg\|\mathcal{D}(C(t_n)) - \mathcal{D}\bigg(\sum_{d\in I} V_d\bigg)\bigg\| &\le \limsup_{n\to\infty} \mathrm{P}\bigg(\sum_{d\in I} V_d < C(t_n)\bigg)\\ &\le \limsup_{n\to\infty} \mathrm{E}(C(t_n)) - \lim_{n\to\infty} \mathrm{E}\bigg(\sum_{d\in I} V_d\bigg)\\ &= \exp(-x) - \exp(-x)\\ &= 0.\end{aligned}$$

7. Local stacks and local alignment

The reader who is familiar with local alignment of biological sequences, that being either amino acid sequences (proteins) or DNA sequences, will have noticed a clear similarity between the results obtained in this paper and results obtained for local alignment. Among the many papers on that subject we refer to the theoretical papers [5], [6], [9], and [19] and the more applied papers [13] and [20]. The result that comes closest to Theorem 1 is the one obtained by Dembo *et al.* [9, Theorem 1] about gapless, local alignment using a general score function.

By reformulating Theorem 1 we see that, for appropriate ts (of order $\log n$),

$$-\log \mathrm{P}(\mathcal{M}_n \le t) \simeq K^* n \exp(-\theta^* t). \tag{19}$$

We find this formulation convenient for comparison with the local alignment results.

For gapless, local alignment we have, in addition to the sequence $(X_k)_{k\ge1}$, an independent sequence $(Y_k)_{k\ge1}$ of i.i.d. variables, and the maximal local similarity between two contiguous parts is defined as

$$\widetilde{\mathcal{M}}_n = \max_{i,j,\delta} \sum_{k=1}^{\delta} f(X_{i+k}, Y_{j+k}),$$

with the maximum taken over $i, j, \delta \ge 0$ such that $i+\delta, j+\delta \le n$. A consequence of [9, Theorem 1] is that, if $\mathrm{E}(f(X_1, Y_1)) < 0$, then, for appropriate ts,

$$-\log \mathrm{P}(\widetilde{\mathcal{M}}_n \le t) \simeq K' n^2 \exp(-\theta' t), \tag{20}$$

where $\theta' > 0$ is the solution to $\mathrm{E}(\exp(\theta f(X_1, Y_1))) = 1$. Representations of the constant K' can be found in [14]; see also [9, Equation (1.2)]. The major assumption for (20) to hold is [9, Condition (E')], which gives a restriction on the distribution of the sequences in relation to the score function used – in addition to requiring a negative score on average. The major assumption in the present paper for (19) to hold is (3), which essentially asks for a sufficiently

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

fast rate of decay of the loop-penalty function g; see [10] and Appendix B. In particular, taking $g \equiv 0$, the condition is violated. However, it is easy to see that $g \equiv 0$ gives a setup very similar to aligning two i.i.d. sequences. When $g \equiv 0$ it is possible to show that

$$-\log \mathrm{P}(\mathcal{M}_n \leq t) \simeq K' \frac{n^2}{2} \exp(-\theta^* t), \tag{21}$$

where K' is the same constant that occurs when aligning two independent sequences with the same distribution as $(X_k)_{k\geq 1}$. This holds if [9, Condition (E')] is fulfilled. We will not give a proof of this, but we may go through the proof in [9] and verify that it carries over almost verbatim. The major difference is that we consider only an upper triangular score matrix, which explains the occurrence of $n^2/2$ in (21) as compared to n^2 in (20).

Summing up, the theory for gapless, local alignment of two random sequences pretty much carries over to provide results for the maximal stack score for a single random sequence when the loop is not penalized ($g \equiv 0$). The present paper treats the case when g decreases sufficiently fast so that (3) holds. It is an open problem to deal with the case when (3) is violated but $g(n) \to -\infty$ as $n \to \infty$.

8. Concluding remarks

We have in this paper provided an affirmative answer to the conjecture stated in [21] in the, quite restrictive, case where we only consider stack RNA structures. Thus, no internal loops, bulges, or multibranch loops are allowed. We have, however, been able to generalize the setup in the direction of allowing for a completely general hairpin-loop penalty function by relying on the results developed in [10].

Applying techniques similar to those in [19], it is expected that the results of the present paper could be generalized to allow for stem-loop structures with a finite or suitably controlled number of internal loops or bulges in the stack. As in [19] such a generalization should affect the constant K^* only (not θ^*).

To allow for general stem-loop or even general secondary structures in the computation of the score as in [21] is, however, a challenge of a substantially different nature. As for local alignments we must expect that strong laws as given in [21] are easier to obtain than limit distributions, and it seems that more sophisticated methods are needed in order to prove distributional limit results for general structures.

Appendix A. Mean value inequalities and Laplace transforms

We show in this appendix some general, useful mean value inequalities that are needed in the paper.

Consider two random variables X and Y taking values in a set E and let $f : E \times E \to \mathbb{R}$ be a given function. Let the distribution of X be π_1 and the distribution of Y be π_2 and let $\pi = \pi_1 \otimes \pi_2$. For the derivations presented in this appendix, we do not need to require that E is finite, but only that the Laplace transform

$$\varphi(\theta) = \mathrm{E}(\exp(\theta f(X, Y))) = \int \exp(\theta f(x, y))\pi(\mathrm{d}x, \mathrm{d}y)$$

of the distribution of $f(X, Y)$ exists (is less than ∞) for all $\theta > 0$, that $\mu = \mathrm{E}(f(X, Y)) < 0$, and, furthermore, that $f(X, Y)$ takes positive values with positive probability. In this case,

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

$\varphi(\theta) \to \infty$ as $\theta \to \infty$, and, since $\partial_\theta \varphi(0) = \mu < 0$, there is a unique solution $\theta^* > 0$ to $\varphi(\theta) = 1$ due to convexity of φ. We define the measure π^* by

$$\frac{\mathrm{d}\pi^*}{\mathrm{d}\pi}(x, y) = \exp(\theta^* f(x, y))$$

and let π_1^* and π_2^* denote the marginals of π^*.

Under π^*, the mean

$$\mu^* = \int f(x, y)\pi^*(\mathrm{d}x, \mathrm{d}y) = \int f(x, y)\exp(\theta^* f(x, y))\pi(\mathrm{d}x, \mathrm{d}y) = \partial_\theta \varphi(\theta^*)$$

is positive, and we ask how this mean relates to the mean of f under $\pi_1^* \otimes \pi_2^*$ as well as under $\pi_1^* \otimes \pi_2$ or $\pi_1 \otimes \pi_2^*$.

Introducing the Laplace transform

$$\varphi^*(\theta) = \int \exp(\theta(f(x, z) + f(w, y) - f(x, y) - f(w, z)))\pi^* \otimes \pi^*(\mathrm{d}x, \mathrm{d}y, \mathrm{d}w, \mathrm{d}z),$$

we see that $\varphi^*(0) = \varphi^*(\theta^*) = 1$, and with

$$\hat{\mu}^* = \int f(x, y)\pi_1^* \otimes \pi_2^*(\mathrm{d}x, \mathrm{d}y)$$

we obtain $\partial_\theta \varphi^*(0) = 2\hat{\mu}^* - 2\mu^*$. Hence, by the convexity of φ^* we obtain $\hat{\mu}^* \leq \mu^*$. If

$$\pi \otimes \pi(\{(x, y, z, w) \mid f(x, z) + f(w, y) \neq f(x, y) + f(w, z)\}) > 0, \tag{22}$$

the Laplace transform φ^* is strictly convex implying that $\hat{\mu}^* < \mu^*$.

Likewise, we can consider the Laplace transform

$$\tilde{\varphi}^*(\theta) = \int \exp(\theta(f(x, z) - f(x, y)))\pi^* \otimes \pi_2(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z),$$

for which $\tilde{\varphi}^*(0) = \tilde{\varphi}^*(\theta^*) = 1$, and with

$$\tilde{\mu}^* = \int f(x, y)\pi_1^* \otimes \pi_2(\mathrm{d}x, \mathrm{d}y),$$

we have $\partial_\theta \tilde{\varphi}^*(0) = \tilde{\mu}^* - \mu^*$. So, if

$$\pi \otimes \pi_2(\{(x, y, z) \mid f(x, z) \neq f(x, y)\}) > 0, \tag{23}$$

then the Laplace transform $\tilde{\varphi}^*$ is strictly convex; hence, $\tilde{\mu}^* < \mu^*$.

We collect these observations into the following lemma for the case in which E is finite.

Lemma 6. *If E is finite, if $\pi_1(x), \pi_2(x) > 0$ for all $x \in E$, and f is not of the form*

$$f(x, y) = f_1(x) + f_2(y) \tag{24}$$

for some $f_1, f_2 : E \to \mathbb{R}$, then

$$\max\{\pi_1^* \otimes \pi_2^*(f), \pi_1^* \otimes \pi_2(f), \pi_1 \otimes \pi_2^*(f)\} < \pi^*(f).$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

Proof. Since $\pi_1(x) > 0$ and $\pi_2(x) > 0$ for all $x \in E$, condition (22) implies that $\pi_1^* \otimes \pi_2^*(f) < \pi^*(f)$ is equivalent to the existence of $x, y, z, w \in E$ such that

$$f(x, z) + f(w, y) \neq f(x, y) + f(w, z).$$

We show by contradiction that if f is *not* of the form (24), then there exist such $x, y, z, w \in E$. Therefore, suppose that, for all $x, y, z, w \in E$, we have

$$f(x, z) + f(w, y) = f(x, y) + f(w, z),$$

then we can fix some $w_0, z_0 \in E$ such that, for all $x, y \in E$, we have

$$f(x, y) = f(x, z_0) - f(w_0, z_0) + f(w_0, y) = f_1(x) + f_2(y),$$

with, for example, $f_1(x) = f(x, z_0) - f(w_0, z_0)$ and $f_2(y) = f(w_0, y)$. This contradicts the assumption that f does not take the form (24).

A similar argument based on (23) instead of (22) implies that $\pi_1^* \otimes \pi_2(f) < \pi^*(f)$, and analogously $\pi_1 \otimes \pi_2^*(f) < \pi^*(f)$.

Note that if $f(x, y) = f_1(x) + f_2(y)$ for some f_1 and f_2 then $\pi^* = \pi_1^* \otimes \pi_2^*$ and the conclusion of Lemma 6 does not hold.

Appendix B. Reflections of random walks

We present in this appendix a summary of the most important constructions and results from [10] adapted to the setup of the present paper.

With the notation as introduced in Section 3 we define the random walk $(S_n)_{n\geq 0}$ by

$$S_n = \sum_{k=1}^{n} f(X_{-k}, X_k)$$

and if $h : \mathbb{N}_0 \to (-\infty, 0]$ is any given function we define the *reflection* of the random walk at the barrier h recursively by $T_0 = h(0)$ and, for $n \geq 1$,

$$T_n = \max\{T_{n-1} + f(X_{-n}, X_n), h(n)\}. \tag{25}$$

The process $(T_n)_{n\geq 0}$ can be expressed as

$$T_n = S_n + \max_{0\leq k\leq n} \{h(k) - S_k\},$$

which follows by verifying that the right-hand side fulfills the recursion (25) and is equal to $h(0)$ for $n = 0$.

In [10] it is (implicitly) assumed that $h(0) = 0$ to make the reflected process start at 0, but in this paper we allow for $h(0) < 0$. Since $T_n - h(0)$ is the reflection of the random walk at the barrier given by $h(n) - h(0)$ for $n \geq 0$, which is 0 for $n = 0$, the results in [10] are easily seen to generalize allowing $h(0) < 0$.

To formulate the main results from [10] we need to introduce the tilted measure P^*. Recall the definition of $\theta^* > 0$, assuming $\mu < 0$, as the unique solution to the equation $\varphi(\theta) = 1$ where $\varphi(\theta) = \mathrm{E}(\exp(\theta f(X_{-1}, X_1)))$. Let π^* denote the probability measure on $E \times E$ given by

$$\pi^*(x, y) = \exp(\theta^* f(x, y))\pi(x)\pi(y).$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

Then P^* is a probability measure such that $(X_{-n}, X_n)_{n\geq 1}$ forms an i.i.d. sequence under P^* with the distribution of (X_{-1}, X_1) being π^*. The mean of $f(X_{-n}, X_n)$ under P^* is

$$\pi^*(f) = \sum_{x,y} f(x, y)\exp(\theta^* f(x, y))\pi(x)\pi(y) = \partial_\theta \varphi(\theta^*) > 0.$$

The probability measure P^* is an example of an exponential change of the measure P, which is defined on the same probability space and related to P as follows. With $\mathcal{F}_n$ the σ-algebra generated by $X_{-n}, \dots, X_n$, the restriction of P^* to $\mathcal{F}_n$ has Radon–Nikodym derivative $\exp(\theta^* S_n)$ with respect to P restricted to $\mathcal{F}_n$, and $\exp(-\theta^* S_n)$ is the Radon–Nikodym derivative the other way around. As a consequence, if τ is a stopping time with respect to the filtration $(\mathcal{F}_n)_{n\geq 1}$ then, for any event $A \in \mathcal{F}_\tau$ with $A \subseteq (\tau < \infty)$, it holds that

$$\mathrm{P}(A) = \mathrm{E}^*(\exp(-\theta^* S_\tau); A), \tag{26}$$

$$\mathrm{P}^*(A) = \mathrm{E}(\exp(\theta^* S_\tau); A); \tag{27}$$

see [7, Theorem XIII.3.2]. Here, E^* denotes expectation under P^*.

If we define

$$M = \sup_{n\geq 0} T_n \quad \text{and} \quad D = \sup_{n\geq 0}\{h(n) - S_n\},$$

the first half of [10, Theorem 2.1] reads as follows.

Theorem 3. *When $\mu < 0$ it holds that*

$$\mathrm{P}(M > u) \leq \exp(-\theta^* u)\,\mathrm{E}^*(\exp(\theta^* D))$$

and $\mathrm{P}(M < \infty) = 1$ *if and only if*

$$\mathrm{E}^*(\exp(\theta^* D)) < \infty.$$

We can elaborate a little on the second half of [10, Theorem 2.1] and provide not only a bound on $\mathrm{E}^*(\exp(\theta^* D))$ but another formula. By partial integration we find that

$$\mathrm{E}^*(\exp(\theta^* D)) = \int_{-\infty}^{\infty} \theta^* \exp(\theta^* u)\,\mathrm{P}^*(D > u)\,\mathrm{d}u.$$

With $\tau(u) = \inf\{n \geq 0 \mid h(n) - S_n > u\}$ for $u \in \mathbb{R}$, (27) then implies that

$$\mathrm{P}^*(D > u) = \mathrm{P}^*(\tau(u) < \infty) = \mathrm{E}(\exp(\theta^* S_{\tau(u)}); \tau(u) < \infty).$$

With

$$D_n = \max_{0\leq k\leq n}\{h(k) - S_k\},$$

it follows that if $D_n - D_{n-1} > 0$ then $S_n + D_n = h(n)$. With $D_{-1} = -\infty$, we find that $\tau(u) = n$ if and only if $D_{n-1} < u \leq D_n$ and, consequently,

$$\begin{aligned}
\mathrm{E}^*(\exp(\theta^* D)) &= \sum_{n=0}^{\infty} \mathrm{E}\biggl(\exp(\theta^* S_n)\int_{D_{n-1}}^{D_n} \theta^* \exp(\theta^* u)\,\mathrm{d}u\biggr) \\
&= \sum_{n=0}^{\infty} \mathrm{E}(\exp(\theta^* S_n)[\exp(\theta^* D_n) - \exp(\theta^* D_{n-1})]) \\
&= \sum_{n=0}^{\infty} \exp(\theta^* h(n))\,\mathrm{E}(1 - \exp(\theta^*(D_{n-1} - D_n))),
\end{aligned} \tag{28}$$

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

with the third equality following from the fact that $S_n + D_n = h(n)$ whenever the integral is nonzero, i.e. whenever $D_n - D_{n-1} > 0$. From this expression, we derive the useful upper bound

$$\mathrm{E}^*(\exp(\theta^* D)) \leq \sum_{n=0}^{\infty} \exp(\theta^* h(n)), \tag{29}$$

because the second factor in (28) is less than or equal to 1.

To formulate the second result from [10] we introduce the stopping time $\tau_+ = \inf\{n \geq 0 \mid S_n > 0\}$, which is finite P^*-a.s., and we define B to be a positive random variable that (under P^*) has distribution given by

$$\mathrm{P}^*(B \leq x) = \frac{1}{\mathrm{E}^*(S_{\tau_+})} \int_0^x \mathrm{P}^*(S_{\tau_+} > y)\,\mathrm{d}y, \qquad x \geq 0.$$

Since we work under the general assumption that the distribution of $f(X_{-1}, X_1)$ is *not* concentrated on a lattice, [10, Theorem 2.3] can be formulated as follows.

Theorem 4. *When* $\mu < 0$ *and* $\mathrm{P}(M < \infty) = 1$, *or equivalently* $\mathrm{E}^*(\exp(\theta^* D)) < \infty$, *it holds that*

$$\mathrm{P}(M > u) \sim \exp(-\theta^* u)\,\mathrm{E}^*(\exp(\theta^* D))\,\mathrm{E}^*(\exp(-\theta^* B))$$

as $u \to \infty$.

References

[1] Altschul, S. *et al.* (1990). Basic local alignment search tool. *J. Molec. Biol.* **215,** 403–410.

[2] Altschul, S. F. *et al.* (1997). Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

[3] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17,** 9–25.

[4] Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5,** 403–434.

[5] Arratia, R., Gordon, L. and Waterman, M. (1986). An extreme value theory for sequence matching. *Ann. Statist.* **14,** 971–993.

[6] Arratia, R., Gordon, L. and Waterman, M. S. (1990). The Erdős–Rényi law in distribution, for coin tossing and sequence matching. *Ann. Statist.* **18,** 539–570.

[7] Asmussen, S. (2003). *Applied Probability and Queues* (Appl. Math. **51**), 2nd edn. Springer, New York.

[8] Dembo, A., Karlin, S. and Zeitouni, O. (1994). Critical phenomena for sequence matching with scoring. *Ann. Prob.* **22,** 1993–2021.

[9] Dembo, A., Karlin, S. and Zeitouni, O. (1994). Limit distribution of maximal non-aligned two-sequence segmental score. *Ann. Prob.* **22,** 2022–2039.

[10] Hansen, N. R. (2006). The maximum of a random walk reflected at a general barrier. *Ann. Appl. Prob.* **16,** 15–29.

[11] Hofacker, I. L., Schuster, P. and Stadler, P. F. (1998). Combinatorics of RNA secondary structures. *Discrete Appl. Math.* **88,** 207–237.

[12] Hofacker, I. L. *et al.* (1994). Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125,** 167–188.

[13] Karlin, S. and Altschul, S. F. (1990). Methods for assessing the statistical significance of molecular features by using general scoring schemes. *Proc. Nat. Acad. Sci.* **87,** 2264–2268.

[14] Karlin, S. and Dembo, A. (1992). Limit distributions of maximal segmental score among Markov-dependent partial sums. *Adv. Appl. Prob.* **24,** 113–140.

[15] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Springer, Berlin.

[16] Mathews, D. H., Sabina, J., Zuker, M. and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Molec. Biol.* **288,** 911–940.

[17] Reinert, G. and Schbath, S. (1998). Compound Poisson and Poisson process approximations for occurrences of multiple words in Markov chains. *J. Comput. Biol.* **5,** 223–253.

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press

[18] Siegmund, D. (1985). *Sequential Analysis*. Springer, New York.

[19] Siegmund, D. and Yakir, B. (2000). Approximate p-values for local sequence alignments. *Ann. Statist.* **28,** 657–680.

[20] Waterman, M. and Vingron, M. (1994). Rapid and accurate estimates of statistical significance for sequence data base searches. *Proc. Nat. Acad. Sci.* **91,** 4625–4628.

[21] Xiong, M. and Waterman, M. S. (1997). A phase transition for the minimum free energy of secondary structures of a random RNA. *Adv. Appl. Math.* **18,** 111–132.

[22] Zuker, M., Mathews, D. and Turner, D. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology*, Kluwer, Dordrecht, pp. 11–43.

https://doi.org/10.1239/aap/1189518638 Published online by Cambridge University Press