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 i.i.d. 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 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 stacks; loop penalty; Poisson approximations; 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 \{a, c, g, 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 [21, 16, 11]. The objective is to maximise 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 [20]. Letting \(M_n\) denote the maximal score for a sequence of \(n\) i.i.d. random variables from the RNA alphabet \{a, c, g, u\} Xiong and Waterman find in [20] the proper scaling of \(M_n\) to obtain strong limit results. More precisely they show 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 [20] that for parameters in the logarithmic phase the normalised, maximal score, \(\theta^* M_n - \log(K^*n)\) for some \(\theta^*, K^* > 0\), 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 report 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 any representation and there is no theoretical results that

∗ Postal address: Department of Applied Mathematics and Statistics, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, richard@math.ku.dk
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. [18].

We offer a solution of a more modest problem for RNA structure than the conjecture in [20]. We will restrict our attention to the maximal score over stacks with a single hairpin loop, that is, no internal loops, bulges or multibranch loops are allowed in the structure – see [12] 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 [20] in our more restrictive setup. In one respect we are, however, capable of being more general than in [20], and that is in the choice of penalty function on the length of the hairpin loop. Where Xiong and Waterman in [20] consider 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 $\mathcal{K}^*$, and we investigate through several examples how the choice of penalty function influences this constant.

If one chooses not to penalise 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.

In the appendix we provide some general inequalities for the expectation of random variables of the form $f(X, Y)$ under various distributional assumptions on $(X, Y)$.

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 with $g(0) = 0$. We define, for $i, j$ and $\delta \geq 0$ satisfying $1 \leq i - \delta$, $i \leq j + 1$, and $j + \delta \leq n$, the random variables

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

Then $S_{i,j}^{\delta} + g(j - i + 1)$ is the (loop-length penalised) score of the stack/hairpin-loop structure as given by $(i, j, \delta)$, 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

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

as the maximal, penalised score.
Local stack scores with general loop penalty

\[
\begin{array}{cccc}
X_1 & X_2 & X_3 & X_4 & X_5 \\
X_6 & X_7 & X_8 & X_9 & X_{10} \\
X_{11} & X_{12} & X_{13} & & \\
\end{array}
\]

**Figure 1:** An graphical illustration of a stack/hairpin-loop structure for the random variables \(X_1, \ldots, X_{13}\). 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 \(X_5, \ldots, X_9\) is 5.

We define a matrix \((T_{i,j})_{i,j \geq 1}\) by

\[
T_{i,j} = \begin{cases} 
T_{i+1,j-1} + f(X_i, X_j), g(j - i + 1) & \text{for } j > i \\
0 & \text{for } j < i \\
T_{i,i} & \text{for } j = i.
\end{cases}
\]

and recursively by

\[
T_{i,j} = \max \left\{ T_{i,j} + 1, T_{i+1,j-1} + f(X_i, X_j), g(j - i + 1) \right\}
\]

(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

\[
T_{i,j} = \max \left\{ S_{i',j'} + g(j' - i' + 1) \right\}
\]

(2)

by verifying that the r.h.s. fulfills the recursion.

If we introduce the set of upper triangular indices

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

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 \(\mathcal{H}_n^0\) by \((i, j) \sim (i', j')\) if \(i' + j' = i + j\) and we can write \(I = \mathcal{H}_n^0/\sim\). 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.

Introducing

\[
\mathcal{M}_n^d = \max_{(i,j) \in d} T_{i,j}
\]

as the maximum of the \(T_{i,j}\)-values along the diagonal \(d\) we find, due to (2) and the fact that the set of diagonals \(I\) forms a partition of \(\mathcal{H}_n^0\), that

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

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

\[
C_n(t) = \sum_{d \in I} 1(\mathcal{M}_n^d > t),
\]

(4)

we get that

\[
(\mathcal{M}_n \leq t) = (C_n(t) = 0).
\]
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 \(\pi\) denote the distribution of \(X_1\) and assume (w.l.o.g.) that \(\pi(x) > 0\) for all \(x \in E\). We will assume that

\[ f(x, y) > 0 \]

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 \(\{f(x, y) | x, y \in E\}\) is not contained in a set of the form \(\delta \mathbb{Z}\) for some \(\delta > 0\).

We denote by

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

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 \(C^\infty\)-function and \(\mu = \partial_\theta \varphi(0)\). If \(\mu < 0\) there is a positive solution, \(\theta^*\), to the equation \(\varphi(\theta) = 1\) since \(\varphi(\theta) \to \infty\) for \(\theta \to \infty\) due to (5). It is unique due to convexity.

Introduce two stochastic processes by the recursive definitions:

\[ T^0_n = \max\{T^0_{n-1} + f(X_{-n}, X_n), g(2n)\}, \quad T^0_0 = 0, \]

and

\[ T^1_n = \max\{T^1_{n-1} + f(X_{-n}, X_n), g(2n + 1)\}, \quad T^1_0 = g(1). \]

We note that for \(d \in I_i, i = 0, 1\), the diagonal \((T^i)_{(i,j) \in d}\) in the \(T^i\)-matrix has the same distribution as (a finite part of) the process \((T^i_n)_{n \geq 0}\). The processes \((T^i_n)_{n \geq 0}, i = 0, 1\), are both random walks reflected in a general barrier, see [10], with the barrier being given by \(g\) evaluated in either the even or the odd integers.

By Theorem 2.2 in [10] it follows that if

\[ M^i := \sup_{n \geq 0} T^i_n < \infty \quad \text{a.s.} \]

for \(i = 0, 1\) then there are constants \(K^*_0\) and \(K^*_1\) such that

\[ P(M^i > x) \sim K^*_i \exp(-\theta^* x) \]

for \(x \to \infty\) and \(i = 0, 1\). Define

\[ K^* = K^*_0 + K^*_1. \]

We should note that for (7) to hold it is obviously necessary (but not sufficient) that \(\mu < 0\). With \(g \equiv 0\) we see that \(P(M^i = \infty) = 1\) even when \(\mu < 0\), and therefore (7) can be viewed as a condition on the penalty function \(g\), see also Section 4 and [10].
Theorem 1. Assume that $\mu < 0$, that $\theta^* > 0$ solves $\varphi(\theta) = 1$, that condition (7) is fulfilled, and that $K^*$ is defined by (8). Let, for $x \in \mathbb{R}$,

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

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

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

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

$$\mathbb{P}(M_n \leq t_n) \to \exp(-\exp(-x))$$

for $n \to \infty$.

Remark 1. To give a precise asymptotic distribution of $M_n$ if $f$ takes lattice values, for example integer values, one needs 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 $x_n = \theta^*(t_n - \lfloor t_n \rfloor)$ then

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

for $n \to \infty$. In particular

$$\mathbb{P}(M_n \leq t_n) - \exp(-\exp(-x + x_n)) \to 0$$

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 Theorem 2.2.

4. The constant $K^*$

The value of the constant $K^*$ depends in a complicated way upon $\pi$, $f$ and $g$. We present here the representation of $K^*_\pi$ as given in [10], and some additional formulas for computing the quantities that enter in this representation.

Let $\pi^*$ be the probability measure on $E \times E$ given by

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

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

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

Let $(S_n)_{n \geq 0}$ be the random walk defined by $S_0 = 0$ and for $n \geq 1$

$$S_n = \sum_{k=1}^{n} f(X_{-k}, X_k).$$
Under \( \mathbb{P}^* \) the random walk has positive drift and \( g \) takes negative values, hence

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

is finite \( \mathbb{P}^* \)-a.s. for \( i = 0, 1 \). Likewise

\[
\tau_\pi = \inf\{ n \geq 0 \mid S_n > 0 \}
\]

is finite \( \mathbb{P}^* \)-a.s. Let \( B \) denote a positive random variable that (under \( \mathbb{P}^* \)) has distribution given by

\[
P^*(\tau_\pi < \infty) = \mathbb{E}^*(\exp(-\theta^* B)),
\]

where \( \mathbb{E}^* \) denotes expectation w.r.t. \( \mathbb{P}^* \). Introduce

\[
C_i = \mathbb{E}^*(\exp(\theta^* D^i)) \mathbb{E}^*(\exp(-\theta^* B)) = C_i C.
\]

One should observe that the second factor in (15), \( C \), does not depend upon \( g \) but only upon \( \pi \) and \( f \). We see that

\[
K^* = (C_0 + C_1) C.
\]

There are several ways to represent and compute \( C \). With \( S^+_n = \max\{S_n, 0\} \), Corollary 8.45 in [17] gives that

\[
C = \exp \left\{ -\sum_{n=1}^{\infty} \frac{1}{n} \left[ \mathbb{E}^*(\exp(-\theta^* S^+_n)) \right] \right\}
\]

which is a consequence of the Spitzer-Baxter identities. An integral representation can also be given, see [17], Theorem 8.51 and the subsequent remarks.

Regarding \( C_i \) for \( i = 0, 1 \) we find, by partial integration, that

\[
C_i = \mathbb{E}^*(\exp(\theta^* D^i)) = \int_{-\infty}^{\infty} \theta^* \exp(\theta^* u) \mathbb{P}^*(D^i > u) \, du.
\]

With \( \tau_\pi(u) = \inf\{ n \geq 0 \mid g(2n + i) - S_n > u \} \) for \( u \in \mathbb{R} \) then

\[
\mathbb{P}^*(D^i > u) = \mathbb{P}^*(\tau_\pi(u) < \infty) = \mathbb{E}(\exp(\theta^* S_{\tau_\pi(u)}); \tau_\pi(u) < \infty),
\]

see Theorem XII.3.2 in [7]. Defining

\[
D^i_n = \max_{0 \leq k \leq n} \{ g(2k + i) - S_k \}
\]
it follows that if $D_n^i - D_{n-1}^i > 0$ then $S_n + D_n^i = g(2n + i)$. With $D_{n-1}^i = -\infty$, we find that $\tau_i(u) = n$ if and only if $D_n^i - u \leq D_{n-1}^i$, and consequently

$$C_i = \sum_{n=0}^{\infty} \mathbb{E} \left( \exp(\theta^* S_n) \int_{D_{n-1}^i}^{D_n^i} \theta^* \exp(\theta^* u) du \right)$$

$$= \sum_{n=0}^{\infty} \mathbb{E} \left( \exp(\theta^* S_n) \left[ \exp(\theta^* D_n^i) - \exp(\theta^* D_{n-1}^i) \right] \right)$$

$$= \sum_{n=0}^{\infty} \exp(\theta^* g(2n + i)) \mathbb{E} (1 - \exp(\theta^* (D_n^i - D_{n-1}^i))), \quad (19)$$

with the third equality following from the fact that $S_n + D_n^i = g(2n + i)$ whenever the integral is non-zero, that is, whenever $D_n^i - D_{n-1}^i > 0$. We find the useful upper bound

$$C_i = \mathbb{E}^\star(\exp(\theta^* D^i)) \leq \sum_{n=0}^{\infty} \exp(\theta^* g(2n + i)) \quad (20)$$

because the second factor in (19) is $\leq 1$. This bound can be used to show that $C_i < \infty$ and, as noted above, Theorem 2.1 in [10] then implies that $\mathbb{P}(M^i < \infty) = 1$ so that condition (7) is fulfilled.

Even though the formula (19) doesn’t seem to provide a useful analytic solution, as it seems hard to find the distribution of the differences $D_n^i - D_{n-1}^i$, the formula is suitable for simulations, as we illustrate in the examples below.

For a linear penalty function $g$ we can obtain another, analytically more useful, 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 $\alpha i$) and $D^i$ is thus the maximum of a random walk. We may first note that for the linear penalty function (20) gives that $C_i < \infty$ whenever $\alpha < 0$ and thus in turn that (7) holds. It follows from the Spitzer-Baxter identity, see Theorem VIII.3.2 in [7], that

$$C_i = \exp(\theta^* \alpha i) \exp \left\{ \sum_{n=1}^{\infty} \frac{1}{n} \left[ \mathbb{E}^\star(\exp(\theta^* \tilde{S}_{0,n}^+)) - 1 \right] \right\}. \quad (21)$$

5. Examples

We consider three examples in detail and focus on the value and computation of $K^\star$. More precisely, we focus on the computation of $C_i$ for $i = 0, 1$ since this is a novel problem. Computing $C = \mathbb{E}^\star(\exp(-\theta^* B))$ is in general not straight forward either, but it is a more classical problem, see [7] and [17]. The formula (17) 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.
Table 1: The value of $C$ with $p_0 = 0.5$, computed using (22), decreases as a function of the parameter $p$, but so does $\mu$. Thus increasing $p$ results in a random walk $(S_n)_{n \geq 0}$ with a larger negative drift and a smaller value of $C$.

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

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

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 one also finds that $\theta^* = 1$. Moreover, the simplicity will allow for some rather explicit expressions. In all the examples below we will also take $p_0 = 1/2$. The function $f$ is seen to be non-lattice 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 non-lattice 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 doesn’t seem to be a simpler way of determining which $p$’s and $p_0$’s that 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

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

we find that

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

and

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

This gives

$$C = \exp \left\{ - \sum_{n=1}^{\infty} \frac{1}{p} \left[ F_{n,p}(n(p, p_0)) + \overline{F}_{n,p_0}(n(p, p_0)) \right] \right\} \frac{p \log \frac{p}{p_0} + (1 - p) \log \frac{1 - p}{1 - p_0}}{p \log \frac{p}{p_0} + (1 - p) \log \frac{1 - p}{1 - p_0}}. \quad (22)$$
Example 1. Consider the linear penalty function \( g(n) = \alpha n \), \( \alpha < 0 \), for which we know that (7) holds. Letting

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

we find from (21) that

\[
C_0 = \exp(2\alpha) \exp \left\{ \sum_{n=1}^{\infty} \frac{1}{n} \left[ \exp(2\alpha n) F_{n,p_0}(n'(\alpha, p, p_0)) + \overline{F}_{n,p}(n'(\alpha, p, p_0)) - 1 \right] \right\}.
\]

Using this formula we see on Figure 2 a graph of \( C_0 \) as a function of \( \alpha \) for three different choices of \( p \) (and with \( p_0 = 1/2 \)). The infinite sum is truncated to 1000 terms. A small \( \alpha \) 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 behaviour of \( C_1 \) parallels \( C_0 \).

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

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

\( \alpha < 0 \), \( n_0 \in \mathbb{N}_0 \). Trivially, (20) implies that \( C_i < \infty \) and thus in turn that (7) holds. Taking \( n_0 = 0 \) we get the linear penalty function, but for \( n_0 > 0 \) we get 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 favour small loops in a non-linear way. Figure 3 shows, based
Figure 3: Considering the piecewise linear penalty function \( g(n) = \alpha \max\{n - n_0, 0\} \), \( \alpha < 0 \), \( n_0 \in \mathbb{N} \), we see (left) \( C_0 \) as a function of \( \alpha \) for different choices of \( n_0 \) and with \( p = 2/3 \) and \( p_0 = 1/2 \). We also show examples of sample paths for the process \( (T_n^0)_{n \geq 0} \) (right) for different choices of \( n_0 \).

on simulations of \( D_n^0 \) (under \( \mathbb{P} \)) and the representation (19) of \( C_0 \), the value of \( C_0 \) as a function of \( \alpha \) for \( n_0 = 10, 100, 200 \) and with \( p = 2/3 \) and \( p_0 = 1/2 \). The infinite sum is truncated to 10000 terms, and since all terms in (19) are positive this gives the upper bound

\[
\sum_{n=1}^{\infty} \exp(g(n)) = \frac{\exp(\alpha(10001 - n_0))}{1 - \exp(\alpha)}
\]

on the error due to the truncation. For \( n_0 = 200 \) and \( \alpha = -0.001 \) this upper bound equals 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 (19) 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 \( \alpha \) as for the linear penalty function hold, that is, small values of \( \alpha \) gives the largest value of \( C_0 \). The effect of \( n_0 \) is also clear. The larger \( n_0 \) is, the longer is the unpenalised, initial segment, and the larger the value of \( M_0^0 \) can become. We see that this gives, as one would anticipate, an increase of \( C_0 \) for increasing \( n_0 \).

Example 3. Consider the logarithmic penalty function \( g(n) = \alpha \log n, \alpha < -1 \). To show that (7) holds using (20) we need \( \alpha < -1 \). One can show that if \( \alpha > -1 \) then \( P(M_t^i = \infty) = 1 \), see Example 2.7 in [10]. Figure 4 shows three sample paths for \( \alpha = -2, -1.1 \) and \( \alpha = -0.5 \) (where \( C_i = \infty \)). One 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 (19), as a function of \( \alpha \) for various values of \( p \) and \( p_0 = 1/2 \). 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 10000, say, is not serious – the upper bound is then 3.98. This is due to the slow convergence of the
Local stack scores with general loop penalty

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 (left) $C_0$ as a function of $\alpha$ for different choices of $p$ and $p_0 = 1/2$. We also see examples of sample paths for the process $(T_n^u)_{n \geq 0}$ (right) for different choices of $\alpha$, $p = 2/3$, and $p_0 = 1/2$.

6. Proofs

The proof is an application of Theorem 1 in [3], which in turn is a consequence of the Chen-Stein method. The sum of the indicator variables $1(M_n^d > t_n)$ is unfortunately not directly suitable for an application of that theorem. 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 $d \in I$,

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

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

$$\lim_{n \to \infty} b_n^{-1} \log n = \lim_{n \to \infty} n^{-\varepsilon} b_n = 0 \quad (23)$$
for all \( \varepsilon > 0 \). The band-limitation is used in the present paper as a technical tool with the sole purpose of proving that \( \sum_{d \in I} 1(M_n^d > t_n) \) 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 equals \( \sum_{d \in I} 1(M_n^d > t_n) \). Thus from a practical point of view, one really only needs to compute the values of \( T_{i,j} \) up to the band-limit, and this can be a serious, computational advantage.

We define, for \( d \in I \),
\[
V_d = 1(M_n^d > t_n).
\]
In the framework of Theorem 1 in [3] we need to define a neighbourhood of dependence for the variable \( V_d \), that is, a subset \( B_d \subseteq I \) such that for \( d' \not\in B_d \) we have that \( V_d \) 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 \( d' \not\in B_d \).

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

**Theorem 2.** If
\[
\lambda_n := \sum_{d \in I} \mathbb{E}(V_d) \longrightarrow \lambda,
\]
for \( n \to \infty \), and
\[
\beta_{1,n} = \sum_{d \in I, d' \in B_d} \mathbb{E}(V_d) \mathbb{E}(V_{d'}) \longrightarrow 0,
\]
\[
\beta_{2,n} = \sum_{d, d' \in B_d, d \neq d'} \mathbb{E}(V_d V_{d'}) \longrightarrow 0,
\]
for \( n \to \infty \), then
\[
\left\| \mathcal{D} \left( \sum_{d \in I} V_d \right) - \text{Poi}(\lambda) \right\| \rightarrow 0.
\]
In fact, the bound
\[
\left\| \mathcal{D} \left( \sum_{d \in I} V_d \right) - \text{Poi}(\lambda_n) \right\| \leq 2(\beta_{1,n} + \beta_{2,n})
\]
always holds.

An essential ingredient in verifying (24), (25), and (26) is a change of measure argument. We also applied it to show (18), but we will give a little more details here. With \( \mathcal{F}_n \) the \( \sigma \)-algebra generated by \( X_{-n}, \ldots, X_n \) one sees that the probability measure \( \mathbb{P} \) restricted to \( \mathcal{F}_n \) has Radon-Nikodym derivative \( \exp(-\theta^* S_n) \) w.r.t. \( \mathbb{P}^* \) restricted to \( \mathcal{F}_n \) (and \( \exp(\theta^* S_n) \) is the Radon-Nikodym derivative the other way around). As a consequence, if \( \tau \) is a stopping time w.r.t. 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
\[
\mathbb{P}(A) = \mathbb{E}^* (\exp(-\theta^* S_{\tau}); A),
\]
see [7, Theorem XIII.3.2].
Lemma 1. Under the assumptions given by (23)

\[ \lambda_n = \sum_{d \in I} \mathbb{E}(V_d) \to \exp(-\lambda) \]

for \( n \to \infty \)

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

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

which can easily be verified. Defining

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

we see that

\[ S_{\tau_i(u)} = u + T_{\tau_i(u)} - u + S_{\tau_i(u)} - T_{\tau_i(u)} \]

\[ \geq u - \max_{0 \leq k \leq \tau_i(u)} \{ g(2k + i) - S_k \} \]

\[ \geq u - \sup_{n \geq 0} \{ g(2n + i) - S_n \} = u - D^i \]

hence for \( A \in \mathcal{F}_{\tau_i(u)} \) with \( A \subseteq (\tau_i(u) < \infty) \) we get from (28) that

\[ \mathbb{P}(A) \leq \exp(-\theta^* u) \mathbb{E}^*(\exp(\theta^* D^i); A). \tag{29} \]

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

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

in \( \mathbb{P}^* \)-probability for \( u \to \infty \), see [17], Chapter IX and in particular Lemma 9.13. Consequently

\[ \frac{\tau(t_n)}{b_n} = \frac{\tau(t_n) \log n}{\log n \cdot b_n} \to 0 \]

in \( \mathbb{P}^* \)-probability for \( n \to \infty \) since \( t_n \sim \log n / \theta^* \), and since \( \mathbb{E}^*(\exp(\theta^* D^i)) < \infty \) by assumption we get in particular that

\[ r_i(n) := \mathbb{E}^*(\exp(\theta^* D^i); \tau(t_n) > b_n) \to 0 \]

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

\[ \mathbb{E}(V_d) = \mathbb{P}(M_n^{d,b} > t_n) = \mathbb{P}(M_{b_n}^i > t_n) \]

where

\[ M_{b_n}^i = \max_{0 \leq k \leq b_n} T_n^i. \]

We see that

\[ \mathbb{P}(M^i > t_n) = \mathbb{P}(M_{b_n}^i > t_n) + \mathbb{P}(M_{b_n}^i \leq t_n, \lambda^i > t_n), \]
and since \((M^i_{b_n} \leq t_n, M^i > t_n) = (b_n < \tau_i(t_n) < \infty)\), it follows from (29), and the fact that \(P^*(\tau_i(u) < \infty) = 1\), that

\[
P(M^i_{b_n} \leq u, M^i > u) \leq \exp(-\theta^* t_n)E^*(\exp(\theta^* D_i); \tau_i(t_n) > b_n) = K_1 n^{-1} r_1(n)
\]

where \(K_1 = \exp(-x)/K^*\). Using (29) again we find for any \(d \in I_i\) that

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

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

\[
(n-1)P(M^0 > t_n) + nP(M^1 > t_n) \sim (n-1)K_0^\ast\exp(-\theta^* t_n) + nK_1^\ast\exp(-\theta^* t_n) \to \exp(-x)
\]

for \(n \to \infty\) we see that

\[
r_2(n) := \left| (n-1)P(M^0 > t_n) + nP(M^1 > t_n) - \exp(-x) \right| \to 0
\]

for \(n \to \infty\). Summing up we have

\[
\sum_{d \in I} E(V_d) - \exp(-x) \leq \sum_{d \in I_0} |E(V_d) - P(M^0 > t_n)| + \sum_{d \in I_1} |E(V_d) - P(M^1 > t_n)|
\]

\[
+ \left| (n-1)P(M^0 > t_n) + nP(M^1 > t_n) - \exp(-x) \right|
\]

\[
\leq K_1 r_0(n) + 4b_n K_0^1 n^{-1} + K_1 r_1(n) + 4b_n K_1^2 n^{-1} + r_2(n) \to 0.
\]

for \(n \to \infty\) by (23).

**Lemma 2.** Under the assumptions given by (23)

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

for \(n \to \infty\).

**Proof.** Using the bound

\[
E(V_d) \leq K_2 n^{-1}
\]

for \(d \in I_i\) as found in the previous proof together with the fact that \(|B_d| \leq 8b_n\) we find that

\[
\sum_{d \in I, d' \in B_d} E(V_d)E(V_{d'}) \leq 8b_n (K_0^0 + K_1^1) n^{-2} \to 0
\]

for \(n \to \infty\) by (23).

**Lemma 3.** Under the assumptions given by (23)

\[
\beta_{3,n} = \sum_{d \in I, d' \in B_d, d \neq d'} E(V_d V_{d'}) \to 0
\]

for \(n \to \infty\).
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), that is, 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 \(x = (x_k)_{k \geq 0}\) an infinite sequence of elements from \(E\). Introduce for \(k \in V\)

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

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

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

In order to apply the Azuma-Hoeffding inequality we need to verify that \(s_V\) enjoys a certain Lipschitz property. If \(x = (x_k)_{k \geq 0}\) and \(y = (y_k)_{k \geq 0}\) 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

\[
|s_V(x) - s_V(y)| \leq 16 \max_{x,y \in E} |f(x, y)|. \tag{30}
\]

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

**Lemma 4.** Let \(\tilde{\mathbb{P}}\) 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 \(\xi_V = \mathbb{E}(S_V)\) denote the expectation of \(S_V\) under \(\tilde{\mathbb{P}}\). If \(\xi_V < 0\) it holds that

\[
\tilde{\mathbb{P}}(S_V \geq 0) \leq \exp \left( -\frac{\xi_V^2}{2|V|\eta^2} \right) \tag{31}
\]

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

\[
Z_k = \mathbb{E}(S_V - \xi_V|\mathcal{F}_k).
\]
Then $(Z_k, \mathcal{F}_k)_{k \in V}$ is a mean 0 martingale. 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 $V$, $\mathcal{F}_{-1}$ the trivial $\sigma$-algebra so that $Z_{-1} = \mathbb{E}(S_V - \xi_V | \mathcal{F}_{-1}) = 0$. Then if we for all $k \in V$ with predecessor $k'$ have

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

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

$$\tilde{\mathbb{P}}(Z_k \geq \lambda) \leq \exp \left( -\frac{\lambda^2}{2 \sum_{m \in V, m \leq k} c_m} \right).$$

It is a direct consequence of the independence assumption and the Lipschitz property of $s_V$, as expressed by (30), that

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

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

$$\tilde{\mathbb{P}}(S_V \geq 0) = \tilde{\mathbb{P}}(S_V - \xi_V \geq -\xi_V) = \tilde{\mathbb{P}}(Z_m \geq -\xi_V) \leq \exp \left( -\frac{\xi_V^2}{2|V|\eta^2} \right).$$

**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 $t \geq 0$

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

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

$$V_1 = \{i-\delta, \ldots, i-1, j+1, \ldots, j+\delta \}, \quad \mathcal{E}_1 = \{\{i-k, j+k\} \mid 1 \leq k \leq \delta\}$$

and $(V_2, \mathcal{E}_2)$ be defined likewise using $(i', j', \delta')$. 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 $S_{i,j}^\delta$ and $S_{i',j'}^{\delta'}$. This implies that if we define

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

together with

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

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

$$\mathbb{P}(S_{(i,j)}^\delta > t, S_{(i',j')}^{\delta'} > t) \leq \mathbb{P}(S_2 > 3/4t, S_2' > 3/4t) = \mathbb{P}(S_2 > 3/4t)\mathbb{P}(S_2' > 3/4t) \leq \exp(-3/2\theta^* t).$$
In particular, (32) holds with $\varepsilon = 1/2$. Suppose instead that $|V| \geq t(4||f||_\infty)^{-1}$. With $S_V = s_V((X_k)_{k \geq 0})$ as above
\[
P(S^\delta_{i,j} > t, S^{\delta'}_{i',j'} > t) \leq P(S^\delta_{i,j} > t, S_V \geq 0) + P(S^{\delta'}_{i',j'} > t, S_V \leq 0).
\] (33)
We introduce the probability measure $P^*_{(i,j,\delta)}$ by
\[
dP^*_{(i,j,\delta)} = \exp(\theta^* S^\delta_{i,j}),
\]
and then, considering the first term in (33), we find that
\[
P(S^\delta_{i,j} > t, S_V \geq 0) = E^*_{(i,j,\delta)}(\exp(-\theta^* S^\delta_{i,j}; S^\delta_{i,j} > t, S_V \geq 0)
\leq \exp(-\theta^* t)P^*_{(i,j,\delta)}(S_V \geq 0).
\]
where $E^*_{(i,j,\delta)}$ denotes expectation under $P^*_{(i,j,\delta)}$. It follows that under $P^*_{(i,j,\delta)}$ the distribution of the sequence of variables $(X_i)_{i \geq 0}$ is as follows:
- The variables $(X_k, X_m)$ for $\{k, m\} \in E_1$ with $k < m$ are i.i.d. with distribution $\pi^*$.
- The variables $X_k$ for $k \notin V_1$ are i.i.d. with distribution $\pi$.
- $(X_k)_{k \in V_1}$ and $(X_k)_{k \notin V_1}$ are independent.

We draw two conclusions. First, for any extension of $E_1$ to a perfect matching $E_\infty$ of $N_0$ it holds that $(X_k, X_m)$ for $\{k, m\} \in E_\infty$ 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 $\zeta := \max\{\pi^*_1 \otimes \pi(f), \pi^*_1 \otimes \pi^*_2(f)\} - \pi^*(f)$. Hence
\[
E(S_V) = \xi_V \leq \zeta |V|
\]
where $\zeta < 0$ according to Lemma 6 in Appendix A. The assumptions of Lemma 4 are fulfilled and we find that, since $|V| \geq t(4||f||_\infty)^{-1}$,
\[
P_{(i,j,\delta)}(S_V \geq 0) \leq \exp\left(-\frac{\zeta^2 |V|}{2\eta^2}\right) \leq \exp\left(-\frac{\zeta^2 t}{8||f||_\infty^2}\right) = \exp(-\theta^* \varepsilon t)
\]
where $\varepsilon = \zeta^2(8||f||_\infty^2)^{-1}$. By a similar argument one can deal with the other term in (33), and this completes the proof.

Proof of Lemma 3. For any $d \in I$ there are at most $b_n^2$ elements $(i, j, \delta) \in H_n$ fulfilling that $(i, j) \in d$ and $j - i + 2\delta \leq 2b_n$. Since $V_d$ indicates that for one such $(i, j, \delta)$ we have $S^\delta_{i,j} > t_n$ it follows from Lemma 5 that for $d, d' \in I$ with $d \neq d'$
\[
E(V_d V_{d'}) \leq b_n^4 \exp(-\theta^* (1 + \varepsilon)t_n).
\]
Then, since $|I| = 2n - 1$, $|B_d| \leq 8b_n$, and $t_n$ is given by (9), we get that
\[
\sum_{d \in I, d' \in B_d, d \neq d'} E(V_d V_{d'}) \leq 16b_n n b_n^4 \exp(-\theta^* (1 + \varepsilon)t_n)
\leq \tilde{K} b_n^5 n^{-\varepsilon} \to 0
\]
for $n \to \infty$ due to (23).
Proof of Theorem 1. Note that
\[ \sum_{d \in I} V_d \leq C(t_n). \]
It is then possible to show that \( E(C(t_n)) \to \exp(-x) \) by the same arguments as in the proof of Lemma 1, but it is actually sufficient to verify the easier result that \( \limsup_{n \to \infty} E(C(t_n)) \leq \exp(-x) \). Indeed,
\[
E(C_n(t_n)) = \sum_{d \in I} \mathbb{P}(M'_n > t_n) \leq (n-1)\mathbb{P}(M^0 > t_n) + n\mathbb{P}(M^1 > t_n) \to \exp(-x)
\]
for \( n \to \infty \). Then by the coupling inequality and the fact that the random variables are integer valued
\[
\limsup_{n \to \infty} \left| D(C(t_n)) - D \left( \sum_{d \in I} V_d \right) \right| \leq \limsup_{n \to \infty} E(C(t_n)) - \lim_{n \to \infty} E \left( \sum_{d \in I} V_d \right) \leq \exp(-x) - \exp(-x) = 0.
\]

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 [18], and the more applied papers [13] and [19]. The result that come closest to Theorem 1 is the one obtained by Amir Dembo, Ofer Zeitouni, and Samuel Karlin [9, Theorem 1] about gapless, local alignment using a general score function.

By reformulating Theorem 1 we see that for appropriate \( t \)'s (of order \( \log n \))
\[ -\log \mathbb{P}(M_n \leq t) \simeq K^*n \exp(-\theta^*t). \]
(34)
We find this formulation convenient for comparison with the local alignment results.

For gapless, local alignment we have, in addition to the sequence \( (X_k)_{k \geq 1} \), an independent sequence \( (Y_k)_{k \geq 1} \) of i.i.d. variables, and the maximal local similarity between two contiguous parts is defined as
\[ \overline{M}_n = \max_{i,j,\delta} \sum_{k=1}^{\delta} f(X_{i+k}, Y_{j+k}), \]
with the maximum taken over \( i,j,\delta \geq 0 \) such that \( i+\delta, j+\delta \leq n \). A consequence of Theorem 1 in [9] is that, if \( E(f(X_1, Y_1)) < 0 \), then for appropriate \( t \)'s
\[ -\log \mathbb{P}(\overline{M}_n \leq t) \simeq K'n^2 \exp(-\theta't), \]
(35)
where $\theta' > 0$ is the solution to $E(\exp(\theta f(X_1, Y_1))) = 1$. Representations of the constant $K'$ can be found in [14], see also formula (1.2) in [9]. The major assumption for (35) to hold is the condition (E') in [9], 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 (34) to hold is (7), which essentially asks for a sufficiently fast rate of decay of the loop-penalty function $g$, see [10]. 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 independent, identically distributed i.i.d. sequence. When $g \equiv 0$ it is possible to show that

$$-\log P(M_n \leq t) \simeq K'n^2/2 \exp(-\theta^* t),$$

(36)

where $K'$ is the same constant the occurs when aligning two independent sequences with the same distribution as $(X_k)_{k \geq 1}$. This holds if the condition (E') in [9] is fulfilled. We will not give a proof of this, but one 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 (36) as compared to $n^2$ in (35).

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 penalised ($g \equiv 0$). The present paper treats the case when $g$ decreases sufficiently fast so that (7) holds. It is an open problem to deal with the case when (7) is violated but $g(n) \to -\infty$ for $n \to \infty$.

8. Concluding remarks

We have in this paper provided an affirmative answer to the conjecture stated in [20] 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 generalise the setup in the direction of allowing for a completely general hairpin-loop penalty function by relying on the results developed in [10].

As for local alignments we must expect that strong laws as given in [20] 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. It is the hope of this author that the present paper, despite its shortcomings in generality, has shed a little more light on the distributional behaviour of RNA structure scores for parameters in the logarithmic phase – a subject that has not been given nearly the same attention as local alignment scores.

Appendix A. Mean value inequalities & 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 $\pi_1$ and the distribution of $Y$ be $\pi_2$ 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
\[ \varphi(\theta) = E(\exp(\theta f(X, Y))) = \int \exp(\theta f(x, y))\pi(dx, dy) \]
(37)
of the distribution of $f(X, Y)$ exists (is $< \infty$) for all $\theta > 0$, that $\mu = E(f(X, Y)) < 0$ 
and, furthermore, that $f(X, Y)$ takes positive values with positive probability. In this case, $\varphi(\theta) \to \infty$ for $\theta \to \infty$, and since $\partial_\theta \varphi(0) = \mu < 0$ there is a unique solution $\theta^* > 0$
to $\varphi(\theta) = 1$ due to convexity of $\varphi$. We define the measure $\pi^*$ by
\[ \frac{d\pi^*}{d\pi}(x, y) = \exp(\theta^* f(x, y)). \]
and let $\pi_1^*$ and $\pi_2^*$ denote the marginals of $\pi^*$.

Under $\pi^*$, the mean
\[ \mu^* = \int f(x, y)\pi^*(dx, dy) = \int f(x, y)\exp(\theta^* f(x, y))\pi(dx, dy) = \partial_\theta \varphi(\theta^*) \]
is positive, and we ask how this mean relates to the mean of $f$ under $\pi_1^* \otimes \pi_2^*$ as well as under $\pi_1^* \otimes \pi_2^* \otimes \pi_2^*$.

Introducing the Laplace transform
\[ \varphi^*(\theta) = \int \exp(\theta(f(x, z) + f(w, y) - f(x, y) - f(w, z))) \pi^* \otimes \pi^* \otimes \pi_2^*(dx, dy, dw, dz), \]
we see that $\varphi^*(0) = \varphi^*(\theta^*) = 1$, and with
\[ \hat{\mu}^* = \int f(x, y)\pi_1^* \otimes \pi_2^*(dx, dy) \]
we obtain that $\partial_\theta \varphi^*(0) = 2\hat{\mu}^* - 2\mu^*$. Hence by convexity of $\varphi^*$ we get that $\hat{\mu}^* \leq \mu^*$. If 
\[ \pi \otimes \pi_2((x, y, z, w) \mid f(x, z) + f(w, y) \neq f(x, y) + f(w, z)) > 0, \]
(38)
the Laplace transform $\varphi^*$ is strictly convex implying that $\hat{\mu}^* < \mu^*$.

Likewise, we can consider the Laplace transform
\[ \tilde{\varphi}^*(\theta) = \int \exp(\theta(f(x, z) - f(x, y))) \pi^* \otimes \pi_2^*(dx, dy, dz), \]
for which $\tilde{\varphi}^*(0) = \tilde{\varphi}^*(\theta^*) = 1$, and with
\[ \hat{\mu}^* = \int f(x, y)\pi_1^* \otimes \pi_2^*(dx, dy) \]
we have that $\partial_\theta \tilde{\varphi}^*(0) = \tilde{\mu}^* - \mu^*$. So if 
\[ \pi \otimes \pi_2((x, y, z) \mid f(x, z) \neq f(x, y)) > 0, \]
(39)
the Laplace transform $\tilde{\varphi}^*$ is strictly convex, hence $\tilde{\mu}^* < \mu^*$.

We collect these observations into the following lemma in case $E$ 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

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

(40)

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

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

Proof. Since $\pi_1(x) > 0$ and $\pi_2(x) > 0$ for all $x \in E$, (38) is equivalent to the existence of $x, y, z, w \in E$ such that

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

If this is not the case we can fix some $w_0, z_0 \in E$ such that for all $x, y \in E$

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

with, for example, $f_1(x) = f(x, z_0) - f(w_0, z_0)$ and $f_2(y) = f(w_0, y)$. Since we assume that $f$ does not take the form (40), we see that (38) is fulfilled and $\pi_1^* \otimes \pi_2^*(f) < \pi^*(f)$.

Likewise, if (39) is not fulfilled then $f(x, y)$ does not depend upon $y$, and again due to $f$ not being of the form (40) we conclude that $\pi_1^* \otimes \pi_2(f) < \pi^*(f)$. A similar argument finally gives that $\pi_1 \otimes \pi_2^*(f) < \pi^*(f)$.

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 the lemma does not hold.

References


