The Annals of Probability
1994, Vol. 22, No. 4, 2022–2039

LIMIT DISTRIBUTION OF MAXIMAL NON-ALIGNED TWO-SEQUENCE SEGMENTAL SCORE

BY AMIR DEMBO,[1] SAMUEL KARLIN[2] AND OFER ZEITOUNI[3]

Stanford University, Stanford University and Technion–Israel Institute of Technology

Consider two independent sequences $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$. Suppose that $X_1, \dots, X_n$ are i.i.d. μ_X and $Y_1, \dots, Y_n$ are i.i.d. μ_Y, where μ_X and μ_Y are distributions on finite alphabets Σ_X and Σ_Y, respectively. A score $F: \Sigma_X \times \Sigma_Y \to \mathbb{R}$ is assigned to each pair (X_i, Y_j) and the maximal nonaligned segment score is $M_n = \max_{0 \le i,j \le n-\Delta, \Delta \ge 0} \{\Sigma_{k=1}^{\Delta} F(X_{i+k}, Y_{j+k})\}$. The limit distribution of M_n is derived here when μ_X and μ_Y are not too far apart and F is slightly constrained.

1. Introduction. Our motivation derives from DNA and protein score-based multiple sequence comparisons. Consider two sequences of length n, $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, where the letters X_i take values in a finite alphabet Σ_X and the letters Y_i take values in a finite alphabet Σ_Y. A real-valued score $F(\cdot, \cdot)$ is assigned to each pair of letters (X_i, Y_j). The maximal segment score allowing shifts is

$$M_n = \max_{\substack{0 \le i,j \le n-\Delta \\ \Delta \ge 0}} \left\{ \sum_{k=1}^{\Delta} F(X_{i+k}, Y_{j+k}) \right\}.$$

Suppose the two sequences are independent: $X_1, \dots, X_n$ i.i.d. following the distribution law μ_X and $Y_1, \dots, Y_n$ i.i.d. following the distribution law μ_Y, where μ_X and μ_Y refer to probabilities on Σ_X and Σ_Y, respectively.

Of primary relevance is the case where the expected score per pair is negative and there is positive probability of attaining some positive pair score. Thus, we assume

$$\text{(H)} \qquad E_{\mu_X \times \mu_Y}(F) < 0, \qquad \mu_X \times \mu_Y(F > 0) > 0,$$

in which case $M_n \to \infty$ is the maximum of segmental scores of negative mean. The hypothesis (H) is in force throughout this paper and it is also assumed that μ_X and μ_Y are strictly positive on Σ_X and Σ_Y, respectively.

Received January 1994; revised February 1994.

[1]Partially supported by NSF Grant DMS-92-09712 and by a US–Israel BSF grant.

[2]Partially supported by NIH Grants 5R01GM10452-29, 8R01HG00335-04 and NSF Grant DMS-86-06244.

[3]Partially supported by a US–Israel BSF grant and the fund for promotion of research at the Technion.

AMS 1991 *subject classifications*. Primary 60F10; secondary 60G70.

Key words and phrases. Large deviations, Chen–Stein method, sequence matching, large segmental sums.

It was shown in [8], Theorem 1, that $M_n/\log n$ converges a.s. to a positive finite constant γ^* defined in terms of appropriate relative entropies. Here we address the problem, mentioned in [8], of evaluating limit laws for M_n or, equivalently, for the dual variables $T_y = \inf\{n: M_n > y\}$. These are closely related to Poisson limit laws for the count

$$\overline{W}_y = \sum_{i \le t_y} \sum_{j \le t_y} \sum_{\Delta=1}^{\min\{i,j\}} 1_{\{\sum_{k=1}^{\Delta} F(X_{i+k-\Delta}, Y_{j+k-\Delta}) > y\}},$$

with the proviso that when (i, j, Δ) is counted, then the triplets (i, j, Δ') for $\Delta' > \Delta$ and $(i+k, j+k, \Delta')$ for $\Delta' \ge k \ge 1$ are not counted (the value of t_y is specified in Theorem 1). To state our main result we need some additional notation. Let $d(\cdot\,,\cdot)$ denote the variational norm between the indicated distributions and let $\mathrm{Po}(\lambda)$ denote the Poisson random variable of parameter λ. Let θ^* and α^* denote the conjugate exponent and conjugate measure, respectively, defined in [8]. That is, determine θ^* as the positive constant [unique, by (H)] satisfying

$$E_{\mu_X \times \mu_Y}\left(e^{\theta^* F}\right) = 1$$

and

$$\frac{d\alpha^*}{d(\mu_X \times \mu_Y)} = e^{\theta^* F}.$$

Let $\Sigma = \Sigma_X \times \Sigma_Y$ be the alphabet of letter pairs and let $M_1(\Sigma)$ denote the set of all probability measures on Σ. The relative entropy of $\nu \in M_1(\Sigma)$ with respect to $\mu \in M_1(\Sigma)$, denoted by $H(\nu \mid \mu)$, is given for $\Sigma = \{b_1, \dots, b_N\}$ by the formula

$$H(\nu \mid \mu) = \sum_{i=1}^{N} \nu(b_i) \log \frac{\nu(b_i)}{\mu(b_i)},$$

with $0 \log 0$ interpreted as 0. In addition to (H), we impose throughout the assumption

$$\text{(E}'\text{)} \qquad H(\alpha^* \mid \mu_X \times \mu_Y) > 2 \max\Big(H(\alpha_X^* \mid \mu_X), H(\alpha_Y^* \mid \mu_Y)\Big),$$

where, for any $\nu \in M_1(\Sigma)$, ν_X and ν_Y denote the marginals of ν on Σ_X and Σ_Y, respectively. In particular we shall use μ to denote the product measure $\mu_X \times \mu_Y$. Note that condition (E$'$) requires strict inequality compared to (E$'$) of [8], which permits equality. Although in general, $\gamma^* \le 2/\theta^*$, it is shown in [8], Theorem 4, that under (E$'$), $\gamma^* = 2/\theta^*$ and that, for identical alphabets, (E$'$) holds whenever $\mu_X = \mu_Y$ and $F(x,y) = F(y,x)$ is not of the form $F(x) + F(y)$. It is easy to check that (E$'$) entails $\alpha^* \ne \alpha_X^* \times \alpha_Y^*$. Let

$$\text{(1.1)} \qquad R_n = \max_{\substack{0 \le i \le n-\Delta \\ \Delta \ge 0}} \left\{ \sum_{k=1}^{\Delta} F(X_{i+k}, Y_{i+k}) \right\}$$

be the maximal segment score between two aligned sequences. It is shown in [11], Theorem A (following [10]) that when $F(X,Y)$ is nonlattice, then

$$\lim_{n\to\infty} P\left(R_n - \frac{\log n}{\theta^*} \le x\right) = \exp\bigl(-K^* \exp(-\theta^* x)\bigr), \tag{1.2}$$

whereas if $F(X,Y)$ is a lattice variable, then

$$\lim_{n\to\infty} \exp\bigl(K^* \exp(-\theta^* x_n)\bigr) P\left(R_n - \frac{\log n}{\theta^*} \le x_n\right) = 1$$

for any bounded sequence x_n such that $x_n + \log n/\theta^*$ are lattice points. The constant K^* is determined from fluctuation sum series identities (see, e.g., [11], (1.8) and (1.11)), and examples for which K^* is explicitly computed are given in [11], Section 3.

The analysis of [8] shows that under condition (E′), the constant limits of $M_n/\log n$ and $R_{n^2}/\log n$ are the same (i.e., then $\gamma^* = 2/\theta^*$). Our main result here establishes that the limit distribution of M_n is the same as that of R_{n^2}.

THEOREM 1. *Assume* (E′) *and* (H). *If $F(X,Y)$ is nonlattice, then*

$$\lim_{n\to\infty} P\left(M_n - \frac{2\log n}{\theta^*} \le x\right) = \exp\bigl(-K^* \exp(-\theta^* x)\bigr), \tag{1.3}$$

and if $F(X,Y)$ is a lattice variable, then

$$\lim_{n\to\infty} \exp\bigl(K^* \exp(-\theta^* x_n)\bigr) P\left(M_n - \frac{2\log n}{\theta^*} \le x_n\right) = 1 \tag{1.4}$$

for any bounded sequence x_n such that $x_n + 2\log n/\theta^*$ are lattice points. Moreover, for $t_y = \sqrt{t}e^{\theta^* y/2}$,

$$\lim_{y\to\infty} d\bigl(\overline{W}_y, \mathrm{Po}(tK^*)\bigr) = 0, \tag{1.5}$$

implying that

$$\lim_{y\to\infty} P(T_y \le t_y) = 1 - \exp(-K^* t), \tag{1.6}$$

where if $F(X,Y)$ is a lattice variable, then $y \to \infty$ in (1.5) *and* (1.6) *via lattice points.*

REMARK 1. In deriving Theorem 1 we assume $F(\cdot,\cdot)$ to be finite-valued, although the possibility of $F(x,y) = -\infty$ for some values of (x,y) is easily accommodated (see also the discussion of [8], Theorem 3). Thus, in the special case of $F(x,x) = 1$ and $F(x,y) = -\infty$ for all $x \ne y$ (with $\Sigma_X = \Sigma_Y$), the limit (1.4) corresponds to the limit distribution of the longest segmental match between the two sequences. In this context, condition (H) holds as soon as $|\Sigma_X| > 1$, whereas condition (E′) reduces to

$$\max\left\{\sum_{i\in\Sigma_X}\mu_X(i)\mu_Y(i)\log\mu_Y(i),\ \sum_{i\in\Sigma_X}\mu_X(i)\mu_Y(i)\log\mu_X(i)\right\} < \frac{1}{2}\lambda^*\log\lambda^*, \tag{1.7}$$

where $\lambda^* = e^{-\theta^*} = \Sigma_i\mu_X(i)\mu_Y(i)$ (and in this case, $K^* = 1-\lambda^*$; see [11], Example 2). For this special case, Theorem 1 was proved earlier in [12], Theorem 2.2, encompassing a wide class of *proximal* ψ-mixing stationary sequences (see [12], (2.11), for the technical definition of proximal sequences). It is easy to check that for i.i.d. sequence letters, the proximality condition of [12], (2.11) implies that (1.7) holds. For related results in the context of longest quality match, see [3], [4].

REMARK 2. Theorem 1 putatively extends to the maximal intersequence segment score involving any subset of r out of s independent sequences, of possibly different lengths $n_1,\dots,n_s$ provided (H) applies for each r-subset and there is a unique dominant subset (having the maximal value of θ^*) for which condition (E_λ) of [8], Section 5, holds with strict inequality.

REMARK 3. In [8], Theorem 4, it is shown that $\gamma^* = 2/\theta^*$ if and only if either (E′) holds or

$$H\big(\alpha^* \mid \mu_X\times\mu_Y\big) = 2\max\Big(H\big(\alpha_X^*\mid\mu_X\big), H\big(\alpha_Y^*\mid\mu_Y\big)\Big),$$

in which case $\alpha^* = \alpha_X^*\times\alpha_Y^*$. For example, this situation occurs for identical alphabets when $\mu_X = \mu_Y$ and $F(x,y) = F(x)+F(y)$. In this context, $M_n \le R_n^X + R_n^Y$, where for each fixed n,

$$R_n^X = \max_{\substack{0\le i\le n-\Delta\\ \Delta\ge 0}}\left\{\sum_{k=1}^{\Delta}F(X_{i+k})\right\},\qquad R_n^Y = \max_{\substack{0\le i\le n-\Delta\\ \Delta\ge 0}}\left\{\sum_{k=1}^{\Delta}F(Y_{i+k})\right\}$$

are two i.i.d. random variables. Assuming for simplicity that $F(X)$ is nonlattice, it follows from (1.2) that

$$\begin{aligned}\lim_{n\to\infty}P\left(M_n-\frac{2\log n}{\theta^*}\le x\right) &\ge \lim_{n\to\infty}P\left(\left(R_n^X-\frac{\log n}{\theta^*}\right)+\left(R_n^Y-\frac{\log n}{\theta^*}\right)\le x\right)\\ &= h\Big(K^*\exp\big(-\theta^*x/2\big)\Big),\end{aligned}$$

where

$$\begin{aligned}h(u) &= \int_{-\infty}^{\infty}\exp\Big(-\big(u^2/K^*\big)\exp(\theta^*z)\Big)\,d\Big[\exp\big(-K^*\exp(-\theta^*z)\big)\Big]\\ &= u\int_0^{\infty}\exp\Big(-u\big(t+1/t\big)\Big)\,dt \ge 1.5u\exp(-2.5u).\end{aligned}$$

Since $K^* > 0$, considering $x \to -\infty$, it is clear that (1.3)–(1.6) do not hold in this case.

REMARK 4. Even when (E′) does not hold, M_n may still possess a limiting extremal distribution of type I (with a different constant $1/\theta^* < \gamma^* < 2/\theta^*$), and this might happen even when the set $\mathcal{M}$ of optimal measures as characterized in [8], Theorem 2, is infinite. For example, let $G_Y(y) = \max_x\{F(x,y)\}$ and

$$\overline{R}_n^Y = \max_{\substack{0 \le j \le n-\Delta \\ \Delta \ge 0}} \left\{ \sum_{k=1}^{\Delta} G_Y(Y_{j+k}) \right\}.$$

Suppose that $E_{\mu_Y}(G_Y) < 0$ and let $\overline{\theta}^*$ denote the unique positive solution of $E_{\mu_Y}(e^{\theta G_Y}) = 1$. Then $\overline{R}_n^Y - \log n/\overline{\theta}^*$ possesses a limit distribution of type I (cf. [11] Theorem A). Let $\overline{\Sigma} = \{(x,y)\colon F(x,y) = G_Y(y)\}$ and define $\beta^* \in M_1(\Sigma_Y)$ such that $d\beta^*/d\mu_Y = \exp(\overline{\theta}^* G_Y)$. If

$$(E_Y) \qquad 2H(\beta^* \mid \mu_Y) > \min_{\nu\,:\,\nu(\overline{\Sigma})=1,\,\nu_Y=\beta^*} H(\nu \mid \mu_X \times \mu_Y),$$

then $\gamma^* = 1/\overline{\theta}^*$ (see [8], (1) and (13)). Clearly, $\overline{R}_n^Y \ge M_n$. In Section 3 we provide a specific example for which (E_Y) holds and show that (E_Y) results with

$$(1.8) \qquad \lim_{n\to\infty} P\big(M_n = \overline{R}_n^Y\big) = 1.$$

Consequently, M_n possesses the same limit distribution of type I as does $\overline{R}_n^Y$.

REMARK 5. In comparison with the recent works [1] and [13], we allow for a general score $F(\cdot,\cdot)$, but accomodate neither insertions nor deletions. Note however that in [1] only the growth order of M_n is found, whereas in [13] the Poisson approximation is established under an additional assumption of a limited number of insertions/deletions.

2. Proof of Theorem 1. Because $\{\overline{W}_y \neq 0\} = \{T_y \le t_y\} = \{M_n > y\}$ for $n = [t_y]$, (1.3) and (1.6) are direct consequences of (1.5), whereas (1.4) holds provided (1.5) applies to any bounded $t = t(y)$. Hence, Theorem 1 amounts to proving that (1.5) holds for any bounded $t = t(y)$. We start with an outline of the main steps in proving this result.

The large deviations analysis of [8] allows us to concentrate on segments of length not exceeding $c_1 y$, whose empirical measure is near α^*. Hence, partitioning both sequences into disjoint blocks of size l_y such that $\exp(\theta^* y) \gg l_y \gg y$, the probability $P(\overline{W}_y \neq W_y)$ approaches 0 as $y \to \infty$, where $W_y = \Sigma_{i,j,\xi} I_{i,j,\xi}$, and the indicator $I_{i,j,\xi} = 1$ if there exists a segmental score exceeding y involving the ith block of the X sequence, the jth block of the Y sequence and a relative shift (alignment) ξ between the indices of the X letters and the corresponding Y letters. Adapting the arguments of [11] and [10], we see in Lemma 1 that

$|E[W_y] - tK^*| \to 0$ as $y \to \infty$. Applying the Chen–Stein method, we show that $d(W_y, \mathrm{Po}(tK^*)) \to 0$, from which (1.5) follows. The main task is in bounding the correlation terms $E(I_{i,j,\xi} I_{i',j',\xi'})$, where large deviations estimates are again decisive, and where the condition (E′) and the restriction to an empirical measure near α^* are needed (see Lemma 2).

Turning now to the detailed proof, let $\|\cdot\|$ denote the variational norm between distributions on Σ and let $G_\eta = \{\nu \in M_1(\Sigma): \|\nu - \alpha^*\| < \eta\}$ denote the corresponding open ball of radius $\eta > 0$, centered at α^*. Let $T^i\mathbf{X} = (X_{i+1}, X_{i+2}, \dots)$, $T^j\mathbf{Y} = (Y_{j+1}, Y_{j+2}, \dots)$ and define the empirical measure

$$L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})} = \frac{1}{\Delta} \sum_{k=1}^{\Delta} \delta_{(X_{i+k}, Y_{j+k})}.$$

For $U \in M_1(\Sigma)$, let

$$M_n^U = \max\left\{ \sum_{k=1}^{\Delta} F(X_{i+k}, Y_{j+k}) \colon 0 \le \Delta \le n,\ i, j \le n - \Delta,\ L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})} \in U \right\},$$

that is, M_n^U is the maximal score among segments with letter pairs having empirical measure in the set U. It is shown in [8], Theorem 3, that if U is a closed set such that $\alpha^* \notin U$, then a.s.,

$$\limsup_{n\to\infty} M_n^U / \log n < 2/\theta^*.$$

Let

$$\overline{M}_n^U = \max\left\{ \sum_{k=1}^{\Delta} F(X_{i+k}, Y_{j+k}) \colon 0 \le \Delta \le c_0 \log n,\ i, j \le n - \Delta, L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})} \in U \right\} \tag{2.1}$$

be the maximal score among segments of length not exceeding $c_0 \log n$ and letter pairs having empirical measure in the set U. It follows from [8], Lemma 1, that for c_0 large enough,

$$\sum_{n=1}^{\infty} P(\overline{M}_n^U \ne M_n^U) < \infty.$$

Consequently, for all $\eta > 0$,

$$\lim_{n\to\infty} P(\overline{M}_n^{G_\eta} \ne M_n) = 0. \tag{2.2}$$

In particular, for c_1 large enough and all $\eta > 0$, it suffices to prove (1.5) with the count $\overline{W}_y$ restricted to triplets (i, j, Δ) for which $\Delta \leq c_1 y$ and $L_\Delta^{(T^{i-\Delta}\mathbf{X}, T^{j-\Delta}\mathbf{Y})} \in G_\eta$. Now let $l_y \geq 3c_1 y$ be a sequence of integers such that $\log l_y / y \to 0$ and $y^2/l_y \to 0$ as $y \to \infty$. Set $m_y = t_y/l_y$. Obviously, $m_y \to \infty$. Because $d(\mathrm{Po}(\lambda), \mathrm{Po}(\lambda'))$ $\leq |\lambda - \lambda'|$, we may assume without loss of generality that m_y (and hence t_y) are integers. Partition the sequence $(X_1, \dots, X_{t_y})$ into blocks of l_y letters each, such that the ith block is $X^i = (X_0^i, X_1^i, \dots, X_{l_y-1}^i)$, where $X_k^i = X_{il_y+k+1}$. Similarly, partition the sequence $(Y_1, \dots, Y_{t_y})$ into blocks of l_y letters each. For $j = 0, \dots, m_y - 1$ and $\xi = 0, 1, \dots, l_y - 1$, let $Y^{j,\xi} = (Y_0^{j,\xi}, Y_1^{j,\xi}, \dots, Y_{l_y-1}^{j,\xi})$ denote the ξ-cyclically-shifted jth block, such that $Y_k^{j,\xi} = Y_{jl_y+1+(\xi+k)_{\mathrm{mod}\, l_y}}$. Let

$$W_y = \sum_{i=0}^{m_y-1} \sum_{j=0}^{m_y-1} \sum_{\xi=0}^{l_y-1} I_{i,j,\xi}, \tag{2.3}$$

where

$$I_{i,j,\xi} = \begin{cases} 1, & \text{if } \max\left\{ \displaystyle\sum_{k=r}^{r+\Delta-1} F(X_k^i, Y_k^{j,\xi}) : l_y - \Delta \geq r \geq 0, \right. \\ & \qquad\qquad \left. c_1 y \geq \Delta \geq 0,\ L_\Delta^{i,j,\xi,r} \in G_\eta \right\} > y, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$L_\Delta^{i,j,\xi,r} = \frac{1}{\Delta} \sum_{k=r}^{r+\Delta-1} \delta_{(X_k^i, Y_k^{j,\xi})}.$$

For $k \leq c_1 y$, let $\mathcal{E}_1(k)$ be the event of a score exceeding y in at least one of the segments of length k that cross the block boundaries in either the X sequence or the Y sequence. Similarly, let $\mathcal{E}_2(k)$ be the event of a score exceeding y in at least one of the segments of length k in which the ξ-shift in $Y^{j,\xi}$ causes a gap in the Y letters. It is easy to check that at most $2t_y m_y(k-1)$ segments are contributing to $\mathcal{E}_i(k)$, for $i = 1, 2$, and therefore by the union of events bound,

$$P\left(\bigcup_{k \leq c_1 y} \mathcal{E}_1(k) \bigcup_{k \leq c_1 y} \mathcal{E}_2(k) \right) \leq 2t_y m_y (c_1 y)^2 \sup_{k \geq 1} P\left(\sum_{i=1}^{k} F(X_i, Y_i) > y \right).$$

Because $E[e^{\theta^* F(X,Y)}] = 1$ and independence

$$P\left(\sum_{i=1}^{k} F(X_i, Y_i) > y \right) \leq E\left(\exp\left(\sum_{i=1}^{k} \theta^* F(X_i, Y_i) \right) \right) \exp(-\theta^* y) = \exp(-\theta^* y)$$

and because by definition $t_y = \sqrt{t}\exp(\theta^* y/2)$, we obtain that

$$P\left(\bigcup_{k\le c_1 y} \mathcal{E}_1(k) \bigcup_{k\le c_1 y} \mathcal{E}_2(k)\right) \le \frac{2t(c_1 y)^2}{l_y} \to 0 \quad \text{as } y \to \infty.$$

Let $\mathcal{E}_3(i, j, \xi)$ be the event that there are $\Delta \le r$ and $r + \Delta' \le r' \le l_y$ such that

$$\sum_{k=r-\Delta+1}^{r} F(X_k^i, Y_k^{j,\xi}) > y, \qquad \sum_{k=r'-\Delta'+1}^{r'} F(X_k^i, Y_k^{j,\xi}) > y.$$

Because R_n is monotone in n [see (1.1)], it follows by conditioning on $\{X_k^i, Y_k^{j,\xi}, k \le r\}$ that $P(\mathcal{E}_3(i, j, \xi)) \le P(R_{l_y} > y)^2$. Consequently, by the union of events bound,

$$P\left(\bigcup_{i,j,\xi} \mathcal{E}_3(i, j, \xi)\right) \le m_y^2 l_y P\big(R_{l_y} > y\big)^2 = t\frac{\exp(\theta^* y)}{l_y} P\big(R_{l_y} > y\big)^2.$$

Hence, the next lemma implies that $P(\cup_{i,j,\xi}\mathcal{E}_3(i, j, \xi)) \to 0$ as $y \to \infty$.

LEMMA 1.

$$\lim_{y\to\infty} \frac{\exp(\theta^* y)}{l_y} P\big(R_{l_y} > y\big) = K^*.$$

It is not hard to check that

$$\{\overline{W}_y \ne W_y\} \subset \bigcup_{k\le c_1 y} \mathcal{E}_1(k) \bigcup_{k\le c_1 y} \mathcal{E}_2(k) \bigcup_{i,j,\xi} \mathcal{E}_3(i, j, \xi).$$

Consequently, in order to prove (1.5), it suffices to show that

$$d\big(W_y, \mathrm{Po}(tK^*)\big) \to_{y\to\infty} 0. \tag{2.4}$$

We will return to the proof of (2.4) after completing the proof of Lemma 1.

PROOF OF LEMMA 1. Following [11], divide the realization of $S_n = \Sigma_{i=1}^n F(X_i, Y_i)$ into successive nonnegative excursions:

$$K_0 = 0,$$
$$K_\nu = \min\big\{k\colon k \ge K_{\nu-1} + 1,\ S_k - S_{K_{\nu-1}} < 0\big\}, \qquad \nu = 1, 2, \ldots,$$

with excursion extremes

$$Q_\nu = \max_{K_{\nu-1}\le k < K_\nu} \big(S_k - S_{K_{\nu-1}}\big).$$

Note that Q_ν are i.i.d. random variables, with common distribution function denoted $G(y)$. Thus, $P(R_{K_m} > y) = 1 - [G(y)]^m$. Fix $\delta > 0$ arbitrarily small and define next $m_\pm = \gamma_\pm l_y / E(K_1)$ with $E(K_1) < \infty$ due to $E_\mu(F) < 0$, where $\gamma_+ \geq (1+\delta)$ and $\gamma_- \leq (1-\delta)$ are chosen as the minimal (maximal) values such that m_+ (and m_-, respectively) are integers [as $y \to \infty$, we have $\gamma_+ \to 1+\delta$ and $\gamma_- \to (1-\delta)$]. Using

$$\lim_{y \to \infty} \big(1 - G(y)\big)\exp(\theta^* y) = E(K_1)K^*,$$

which is provided by [11], Lemma A, and the identification of K^* in [11] (below (1.12); see also [10]), one sees that

$$\begin{aligned}\lim_{y\to\infty} \frac{\exp(\theta^* y)}{l_y} P\big(R_{K_{m_+}} > y\big) \\ = \lim_{y\to\infty} \frac{\exp(\theta^* y)}{l_y}\big[1 - G(y)^{\gamma_+ l_y/E(K_1)}\big] = (1+\delta)K^*\end{aligned} \tag{2.5}$$

and

$$\lim_{y\to\infty} \frac{\exp(\theta^* y)}{l_y} P\big(R_{K_{m_-}} > y\big) = (1-\delta)K^*. \tag{2.6}$$

Because R_n is monotone in n,

$$P\big(R_{K_{m_-}} > y\big) - P\big(K_{m_-} > l_y\big) \leq P\big(R_{l_y} > y\big) \leq P\big(R_{K_{m_+}} > y\big) + P\big(K_{m_+} < l_y\big). \tag{2.7}$$

Let $g(\theta) = -\theta + ((1-\delta)/E(K_1)) \log E(\exp(\theta K_1))$. Note that for each m, K_m is a sum of i.i.d. positive random variables. Hence, using Chebycheff's bound,

$$P(K_{m_-} > l_y) \leq \inf_{\theta \geq 0} \Big\{\exp(-\theta l_y)\big(E\exp(\theta K_1)\big)^{m_-}\Big\} \leq \inf_{\theta\geq 0} \exp\big(g(\theta) l_y\big).$$

Note that for $\lambda_0 > 0$ such that $\Lambda(\lambda_0) = \log E(\exp(\lambda_0 F(X_1, Y_1))) < 0$ (λ_0 exists due to the boundedness of F and (H); see [8], proof of Lemma 1), we have

$$P(K_1 > n) \leq P\left(\sum_{i=1}^n F(X_i, Y_i) \geq 0\right) \leq \exp\big(n\Lambda(\lambda_0)\big).$$

Therefore, $g(\theta) < \infty$ for all θ in a small enough neighborhood of 0. It follows that $g'(0) = -\delta < 0$, leading to

$$\frac{\exp(\theta^* y)}{l_y} P(K_{m_-} > l_y) \leq \exp\big(-c(\delta) l_y\big)\exp(\theta^* y) \to_{y\to\infty} 0 \tag{2.8}$$

for some constant $c(\delta) > 0$. A similar computation yields

$$\frac{\exp(\theta^* y)}{l_y} P\big(K_{m_+} < l_y\big) \to_{y\to\infty} 0. \tag{2.9}$$

Substituting (2.5), (2.6), (2.8) and (2.9) into (2.7) and taking $\delta \to 0$ yields the lemma. □

For the objective of proving (2.4), we employ a version of the Chen–Stein method given in [2]. Let $\alpha = (i, j, \xi)$ and let $\mathcal{B}_\alpha = \{(i', j', \xi'): i = i' \text{ or } j = j'\}$ denote the associated neighborhood of dependence. With this definition, note that I_α is independent of $\{I_\gamma: \gamma \notin \mathcal{B}_\alpha\}$. Thus, from [2] (see also [7], inequalities (2.4) and (2.7)), one has

$$d\big(W_y, \mathrm{Po}(tK^*)\big) \le (b_1 + b_2)\frac{(1 - e^{-\lambda_y})}{\lambda_y} + |\lambda_y - tK^*|,$$

where $\lambda_y = E(W_y)$ and

$$b_1 = \sum_\alpha \sum_{\beta \in \mathcal{B}_\alpha} P(I_\alpha = 1)P(I_\beta = 1),$$

$$b_2 = \sum_\alpha \sum_{\substack{\beta \in \mathcal{B}_\alpha \\ \beta \ne \alpha}} P(I_\alpha = 1,\ I_\beta = 1)$$

(in the notations of [2], $b_3 = 0$). Let

$$R_{l_y}^{G_\eta} = \max\left\{ \sum_{k=1}^{\Delta} F(X_{i+k}, Y_{i+k}): 0 \le i \le l_y - \Delta, 0 \le \Delta \le c_1 y,\ L_\Delta^{T^i\mathbf{X}, T^i\mathbf{Y}} \in G_\eta \right\}$$

and $p_y = P(R_{l_y}^{G_\eta} > y)$. Note that for any α, $P(I_\alpha = 1) = p_y$ and $|\mathcal{B}_\alpha| \le 2m_y l_y$. Therefore,

$$\lambda_y = m_y^2 l_y p_y = t\left(\frac{p_y}{l_y}\right)\exp(\theta^* y). \tag{2.10}$$

and

$$b_1 = p_y^2 \sum_\alpha |\mathcal{B}_\alpha| \le 2m_y l_y p_y^2 (m_y^2 l_y) = \frac{2\lambda_y^2}{m_y}.$$

Because $R_{l_y} \ge R_{l_y}^{G_\eta}$, it follows that

$$P(R_{l_y} > y) \ge p_y \ge P(R_{l_y}^{G_\eta} = R_{l_y} \mid R_{l_y} > y) P(R_{l_y} > y).$$

The strong laws of [6], Theorems 1 and 2, imply that $P(R_{l_y}^{G_\eta} = R_{l_y} \mid R_{l_y} > y) \to 1$ for every $\eta > 0$ and, hence by (2.10) and Lemma 1,

$$\lim_{y \to \infty} |\lambda_y - tK^*| = \lim_{y \to \infty} t\left|\left(\frac{p_y}{l_y}\right)\exp(\theta^* y) - K^*\right| = 0 \tag{2.11}$$

[recall that $t = t(y)$ is bounded]. In particular, (2.11) implies that $b_1 \to 0$, and (2.4) thus follows from the next lemma, completing the proof of Theorem 1. □

LEMMA 2. *For all $\eta > 0$ small enough, $b_2 \to 0$ as $y \to \infty$.*

PROOF OF LEMMA 2. Using I_0 to abbreviate $I_{(0,0,0)}$, let $Q_0(y) = \exp(\theta^* y/2) P(I_{(1,0,0)} = 1 \mid I_0 = 1)$, $Q_1(y) = \exp(\theta^* y/2) P(I_{(0,1,0)} = 1 \mid I_0 = 1)$ and $Q_2(y) = \Sigma_{\xi=1}^{l_y - 1} P(I_{(0,0,\xi)} = 1 \mid I_0 = 1)$. By the symmetry of the problem,

$$
\begin{aligned}
b_2 &= \sum_{\alpha} p_y \sum_{\substack{\beta \in \mathcal{B}_{(0,0,0)} \\ \beta \neq (0,0,0)}} P(I_\beta = 1 \mid I_0 = 1) \\
&\le p_y m_y^2 l_y m_y l_y \Bigg[P\big(I_{(1,0,0)} = 1 \mid I_0 = 1\big) + P\big(I_{(0,1,0)} = 1 \mid I_0 = 1\big) \\
&\qquad\qquad + \frac{1}{m_y l_y} \sum_{\xi=1}^{l_y - 1} P\big(I_{(0,0,\xi)} = 1 \mid I_0 = 1\big) \Bigg] \qquad (2.12) \\
&= a_y \big(Q_0(y) + Q_1(y)\big) + \tilde{a}_y Q_2(y),
\end{aligned}
$$

where $a_y = (p_y/l_y) m_y^3 l_y^3 \exp(-\theta^* y/2)$ is such that $|a_y - t^{3/2} K^*| \to 0$ as $y \to \infty$ [see (2.11)] and $\tilde{a}_y = a_y \exp(\theta^* y/2)/l_y m_y$ is such that $|\tilde{a}_y - tK^*| \to 0$ as $y \to \infty$. Proving Lemma 2 thus requires showing that $Q_i(y) \to 0$, $i = 0, 1, 2$, as $y \to \infty$. Due to the symmetric roles played by μ_X and μ_Y, it is enough to consider only $i = 1$ and $i = 2$.

It is now useful to decompose the events I_0, $I_{(0,1,0)}$ and $I_{(0,0,\xi)}$. Thus, let

$$
J_{x,k,\nu} = \left\{ \omega \colon \frac{1}{k} \sum_{j=0}^{k-1} \delta_{(X_{x+j}, Y_{x+l_y+j})} = \nu \in G_\eta,\ kE_\nu(F) > y \right\},
$$

with $x = 1, \dots, l_y - k + 1$, $k \le c_1 y$, and ν ranges over all possible k-types [$\nu \in M_1(\Sigma)$ with $k\nu(i)$ an integer for all $i \in \Sigma$]. Thus, the range of the pair (k, ν) is of cardinality at most $(c_1 y + 1)^{|\Sigma|}$. Similarly, define

$$
J_{x',k',\nu',\xi} = \left\{ \omega \colon \frac{1}{k'} \sum_{j=0}^{k'-1} \delta_{\left(X_{x'+j}, Y_{x'+(\xi+j)_{\mathrm{mod}\, l_y}}\right)} = \nu' \in G_\eta, k' E_{\nu'}(F) > y \right\}
$$

with $x' = 1, \dots, l_y - k' + 1$, $k' \le c_1 y$, ν' ranges over all possible k'-types and $\xi = 0, 1, \dots, l_y - 1$.

Starting with $Q_1(y)$, note that

$$
\begin{aligned}
P\big(I_{(0,1,0)} = 1 \mid I_0 = 1\big) &= P\Bigg(\bigcup_{x,k,\nu} J_{x,k,\nu} \,\Bigg|\, \bigcup_{x',k',\nu'} J_{x',k',\nu',0} \Bigg) \qquad (2.13) \\
&\le \sum_{x,k,\nu} \sum_{x',k',\nu'} P\big(J_{x,k,\nu} \mid J_{x',k',\nu',0}\big).
\end{aligned}
$$

There are two distinct classes of four-tuples $e = (x, x', k, k')$ to consider, $e \in \mathcal{E}_a$ if $[x, x+k-1] \cap [x', x'+k'-1] = \emptyset$ and $e \in \mathcal{E}_b$ otherwise. For $e \in \mathcal{E}_a$,

$$P\big(J_{x,k,\nu} \mid J_{x',k',\nu',0}\big) = P(J_{x,k,\nu}) \le P\big(I_{(0,1,0)} = 1\big) = p_y. \tag{2.14}$$

Because the only connection between the conditioning event and $J_{x,k,\nu}$ is through the X-sequence,

$$\begin{aligned}
&\sup_{e \in \mathcal{E}_b, \nu, \nu'} P\big(J_{x,k,\nu} \mid J_{x',k',\nu',0}\big) \\
&\quad = \sup_{\substack{k,k',\nu,\nu' \\ 1 \le x \le k'}} P\big(J_{x,k,\nu} \mid J_{1,k',\nu',0}\big) \\
&\quad \le \sup_{\substack{(a_1,\dots,a_k) \\ k',k,\nu,1 \le x \le k'}} P\big(J_{x,k,\nu} \mid X_x = a_1, \dots, X_{x+k-1} = a_k\big) \\
&\quad = \sup_{k,\nu} P\left(k^{-1} \sum_{j=1}^{k} \delta_{(X_j, Y_{l_y+j})} = \nu \,\middle|\, k^{-1} \sum_{j=1}^{k} \delta_{X_j} = \nu_X \right) \\
&\quad = \sup_{k,\nu} \frac{P\big(k^{-1} \sum_{j=1}^{k} \delta_{(X_j, Y_j)} = \nu\big)}{P\big(k^{-1} \sum_{j=1}^{k} \delta_{X_j} = \nu_X\big)}.
\end{aligned} \tag{2.15}$$

Using simple combinatorial bounds (see, e.g., [8], (3) and (4)), one sees that

$$\begin{aligned}
&\sup_{k,\nu} \frac{P\Big((k^{-1}) \sum_{j=1}^{k} \delta_{(X_j, Y_j)} = \nu\Big)}{P\Big((k^{-1}) \sum_{j=1}^{k} \delta_{X_j} = \nu_X\Big)} \\
&\quad \le \sup_{k,\nu} (c_1 y + 1)^{|\Sigma|} \exp\Big(-k\big[H(\nu \mid \mu) - H(\nu_X \mid \mu_X)\big]\Big).
\end{aligned} \tag{2.16}$$

By (E′) and the continuity in a of $H(a \mid b)$, for $\eta > 0$ small enough,

$$\beta(\eta) = \inf_{\nu \in G_\eta} \Big\{ H(\nu \mid \mu) - 2\max\big[H(\nu_X \mid \mu_X), H(\nu_Y \mid \mu_Y)\big] \Big\} > 0.$$

Thus, for $\nu \in G_\eta$ such that $kE_\nu(F) > y$, one has that $H(\nu \mid \mu) \ge 2H(\nu_X \mid \mu_X) + \beta(\eta)$, whereas $kH(\nu \mid \mu) \ge \theta^* y$. Hence, using (2.14)–(2.16) and (2.11),

$$\begin{aligned}
Q_1(y) &= \exp(\theta^* y/2) P\big(I_{(0,1,0)} = 1 \mid I_0 = 1\big) \\
&\le \exp(\theta^* y/2) \Big[l_y^2 (c_1 y + 1)^{3|\Sigma|} \exp\big(-\theta^* y/2 - \beta(\eta) y / 2\|F\|_\infty\big) \\
&\qquad + p_y (c_1 y + 1)^{2|\Sigma|} l_y^2 \Big] \to_{y \to \infty} 0.
\end{aligned}$$

It remains to deal with $Q_2(y)$. As in the preceding computation, note that

$$p_y = P\left(\bigcup_{x,k,\nu} J_{x,k,\nu,\xi} \right) = P\left(\bigcup_{x',k',\nu'} J_{x',k',\nu',0} \right)$$

and one has

$$
\begin{aligned}
Q_2(y) &= \sum_{\xi=1}^{l_y-1} P\big(I_{(0,0,\xi)} = 1 \mid I_0 = 1\big) \\
&= \sum_{\xi=1}^{l_y-1} P\Bigg(\bigcup_{x,k,\nu} J_{x,k,\nu,\xi}, \bigcup_{x',k',\nu'} J_{x',k',\nu',0}\Bigg) \Bigg/ p_y \\
&\le 2 \sum_{\substack{\xi,x,k,\nu \\ x',k'\le k,\nu'}} \frac{P(J_{x,k,\nu,\xi}, J_{x',k',\nu',0})}{p_y}.
\end{aligned} \tag{2.17}
$$

For any five-tuple $e = (\xi, x, x', k, k')$, let Δ_X (Δ_Y) denote the set of X_i (Y_i) letters occurring in the definition of $J_{x,k,\nu,\xi}$ that do not occur in the definition of $J_{x',k',\nu',0}$. Three distinct cases are possible:

(a) $|\Delta_X| \vee |\Delta_Y| \ge (1-\eta)k$ (denoted $e \in \mathcal{E}_a$).
(b) $(1-\eta)k \ge |\Delta_X| \vee |\Delta_Y| \ge \delta y$ (denoted $e \in \mathcal{E}_b$).
(c) $|\Delta_X| \vee |\Delta_Y| \le \delta y$ (denoted $e \in \mathcal{E}_c$).

Here, δ is a small fixed constant that depends on η and will be chosen below. We analyze the three cases separately. The argument for $|\Delta_X| > |\Delta_Y|$ being the same as for $|\Delta_X| \le |\Delta_Y|$, we may assume the latter in subsequent computations.

Case (a). To simplify the notations, we assume that ηk is an integer (otherwise replace ηk by its integer part) and let $L_\eta = \sum_{i=1}^{\eta k} \delta_{Y_i}/\eta k$ and $L_{1-\eta} = \sum_{i=\eta k+1}^{k} \delta_{Y_i}/(1-\eta)k$. Note that after relabeling the random variables involved, because $\nu \in G_\eta$, for $\eta \le \frac{1}{2}$,

$$
\begin{aligned}
&P\big(J_{x,k,\nu,\xi} \mid J_{x',k',\nu',0}\big) \\
&\quad\le \sup_{(b_1,b_2,\dots,b_{\eta k})} P\Bigg(\frac{1}{k}\sum_{i=1}^{k} \delta_{Y_i} = \nu_Y \mid Y_1 = b_1, Y_2 = b_2, \dots, Y_{\eta k} = b_{\eta k}\Bigg) \\
&\quad= \sup_{(b_1,b_2,\dots,b_{\eta k})} P\big((1-\eta)L_{1-\eta} + \eta L_\eta = \nu_Y \mid Y_1 = b_1, Y_2 = b_2, \dots, Y_{\eta k} = b_{\eta k}\big) \\
&\quad\le \sup_{\phi \in G_{4\eta}} P(L_{1-\eta} = \phi_Y).
\end{aligned}
$$

With $\alpha_Y^* \ne \mu_Y$, one may find an η small enough such that

$$
\rho(\eta) = \inf_{\phi \in G_{4\eta}} H\big(\phi_Y \mid \mu_Y\big) > 0.
$$

Choosing η at least that small, by the combinatorial upper bound of [9], Lemma 2.1.9,

$$
\sup_{\phi \in G_{4\eta}} P(L_{1-\eta} = \phi_Y) \le \exp\big(-(1-\eta)\rho(\eta)k\big) \le \exp\big(-(1-\eta)\rho(\eta)y/\|F\|_\infty\big)
$$

[recall that $kE_\nu(F) > y$]. Because $p_y \geq P(J_{x',k',\nu',0})$, we are led to the conclusion that, for all $e \in \mathcal{E}_a$,

$$\begin{aligned}\frac{P(J_{x,k,\nu,\xi}, J_{x',k',\nu',0})}{p_y} &\leq P\big(J_{x,k,\nu,\xi} \mid J_{x',k',\nu',0}\big) \\ &\leq \exp\big(-(1-\eta)\rho(\eta)y/\|F\|_\infty\big).\end{aligned} \tag{2.18}$$

Note that in both cases (b) and (c), because the overlap between the sequences involved in the definition of $J_{x,k,\nu,\xi}$ and $J_{x',k',\nu',0}$ is at least of one symbol, whereas $l_y \geq 3c_1 y \geq 3k$, one may relabel the sequences such that $x' = 1$, x may assume both positive and negative values and the modulus operation is omitted from the definition of $J_{x,k,\nu,\xi}$. We will henceforth work with this relabeling without mentioning it further.

Case (b). Let here

$$L_{x,k,\xi} = \frac{1}{k}\sum_{l=x}^{x+k-1} \delta_{X_l, Y_{l+\xi}}$$

and

$$L^{\Delta}_{x,k,\xi} = \frac{1}{k-|\Delta_Y|}\sum_{\substack{l+\xi \in [x+\xi,\\ x+\xi+k-1]\setminus\Delta_Y}} \delta_{X_l, Y_{l+\xi}}.$$

Note that now

$$\begin{aligned}P(J_{x,k,\nu,\xi}, J_{1,k',\nu',0}) &= P(L_{x,k,\xi} = \nu,\ L_{1,k',0} = \nu') \\ &\leq P\big(L_{x,k,\xi} = \nu,\ L^{\Delta}_{x,k,\xi} \notin G_{2\eta}\big) \\ &\quad + \sup_{\phi \in G_\eta} P\big(L_{x,k,\xi} = \phi,\ L_{1,k',0} = \nu',\ L^{\Delta}_{x,k,\xi} \in G_{2\eta}\big) \\ &= A_1 + A_2.\end{aligned}$$

Turning our attention to A_1, note that by combining [9], (2.1.32) and (2.1.34),

$$P\big(L^{\Delta}_{x,k,\xi} = \psi \mid L_{x,k,\xi} = \nu\big) \leq (k+1)^{2(|\Sigma|+1)}\exp\Big(-\big(k-|\Delta_Y|\big)H\big(\psi \mid \nu\big)\Big).$$

Hence, for $\nu \in G_\eta$ such that $kE_\nu(F) > y$,

$$\begin{aligned}A_1/p_y &\leq P\big(L^{\Delta}_{x,k,\xi} \notin G_{2\eta} \mid L_{x,k,\xi} = \nu\big) \\ &\leq (k+1)^{3(|\Sigma|+1)}\exp\Big(-\eta k \inf_{\psi \notin G_{2\eta}} H\big(\psi \mid \nu\big)\Big) \\ &\leq (c_1 y + 1)^{3(|\Sigma|+1)}\exp\big(-y\eta^3/2\|F\|_\infty\big),\end{aligned} \tag{2.19}$$

where we have used in the last inequality the relation (see [9], Exercise 6.2.17)

$$H\big(\psi \mid \phi\big) \geq \|\psi - \phi\|^2/2. \tag{2.20}$$

To evaluate A_2, let L^{Δ_Y} denote the empirical measure of the Y_i letters in the set Δ_Y and note that, denoting $v_\Delta = |\Delta_Y|/k$,

$$\begin{aligned}P\big(L_{x,k,\xi} = \phi,\ L^{\Delta}_{x,k,\xi} \in G_{2\eta} \mid L_{1,k',0} = \nu'\big) &\le (c_1 y+1)^{|\Sigma|} \sup_{\substack{\phi\in G_\eta,\\ \psi\in G_{2\eta}}} P\big(v_\Delta L^{\Delta_Y} + (1-v_\Delta)\psi_Y = \phi_Y\big)\\ &\le (c_1 y+1)^{|\Sigma|} P\big(\|L^{\Delta_Y} - \alpha^*_Y\| \le 3c_1\eta/\delta\big).\end{aligned}$$

Therefore, using again (2.20) and the combinatorial upper bound from [9], Lemma 2.1.9, and choosing $\delta = \delta(\eta)$ not too small such that $3c_1\eta/\delta < \|\alpha^*_Y - \mu_Y\|/2$ (this is always possible for small η because $\alpha^*_Y \ne \mu_Y$), one obtains

$$A_2/p_y \le (c_1 y+1)^{2|\Sigma|}\exp\big(-\delta y\|\alpha^*_Y - \mu_Y\|^2/8\big). \tag{2.21}$$

Note that one may have both η small and $\delta = \delta(\eta)$ small [for example, by choosing $\delta = \delta(\eta) = \sqrt{\eta}$ and taking η small enough]. Combining (2.19) and (2.21), one obtains that for any $e \in \mathcal{E}_b$ and every $\eta > 0$ small enough,

$$\frac{P(J_{x,k,\nu,\xi}, J_{x',k',\nu',0})}{p_y} \le g_1(y)\exp\big(-\kappa(\eta)y\big), \tag{2.22}$$

where $g_1(y)$ is independent of e and of η, $y^{-1}\log g_1(y) \to 0$ with y and $\kappa(\eta) > 0$.

Case (c). Note that because $k \ge k'$ and $|\Delta_X| \le |\Delta_Y| \le \delta y$, necessarily $k - k' \le \delta y$ and $\xi \le 2\delta y$. Let now $Z_i = ((Z_i)_X, (Z_i)_Y)$ denote the following (relabeled) random variables:

$$(Z_i)_X = X_{x-1+i\xi},\qquad (Z_i)_Y = Y_{x-1+(i+1)\xi},\qquad i = 0,1,\dots,\big([k/\xi]-1\big),$$

$$(Z_i)_X = X_{x+(i-[k/\xi])\xi},\qquad (Z_i)_Y = Y_{x+(i-[k/\xi]+1)\xi},\qquad i = [k/\xi],\dots,\big(2[k/\xi]-1\big),$$

and so forth, up to $i = [k/\xi]\xi - 1$. Complete this construction up to $i = k$ in such a way that the empirical measure of $(Z_1,\dots,Z_k)$ is $L_{x,k,\xi}$. Define next the empirical measure

$$L_k = \frac{1}{k}\sum_{i=0}^{k-1} \delta_{Z_i Z_{i+1}} \in M_1(\Sigma^2).$$

For any $\theta \in M_1(\Sigma_X \times \Sigma_Y \times \Sigma_X \times \Sigma_Y)$, let

$$(\theta)_1 = \sum_{\substack{x_2\in\Sigma_X\\ y_2\in\Sigma_Y}} \theta(\cdot,\cdot,x_2,y_2) \in M_1(\Sigma),$$

$$(\theta)_2 = \sum_{\substack{x_1\in\Sigma_X\\ y_1\in\Sigma_Y}} \theta(x_1,y_1,\cdot,\cdot) \in M_1(\Sigma)$$

and

$$(\theta)_{12} = \sum_{\substack{x_1 \in \Sigma_X \\ y_2 \in \Sigma_Y}} \theta(x_1, \cdot, \cdot, y_2) \in M_1(\Sigma).$$

Note that $(L_k)_2 = L_{x,k,\xi}$, $\|(L_k)_1 - L_{x,k,\xi}\| \le 2/k$ and $\|(L_k)_{12} - L_{1,k',0}\| \le (4\xi + 4\delta y)/k \le 12\delta y/k \le 12\delta\|F\|_\infty$. Hence, with $\varepsilon = \eta + 12\delta\|F\|_\infty$, for all large y,

$$\begin{aligned} &P(J_{x,k,\nu,\xi}, J_{1,k',\nu',0}) \\ &\quad \le (c_1 y + 1)^{2|\Sigma|} \sup\nolimits_{\theta_1,\theta_2 \in G_\varepsilon} P\big((L_k)_1 = \theta_1,\ (L_k)_{12} = \theta_2, (L_k)_2 = \nu\big). \end{aligned} \tag{2.23}$$

For any ν, it follows from the Markov structure of the chain $\{(Z_i Z_{i+1})\}_i$ that

$$P(L_k = \nu) \le \exp\Big(-kH\big(\nu \mid (\nu)_1 \times \mu_X \times \mu_Y\big)\Big) \tag{2.24}$$

(see [5], Lemma 3 or [9], Exercise 3.1.21). Using (2.23) and (2.24), one obtains that

$$\begin{aligned} &\frac{P(J_{x,k,\nu,\xi}, J_{x',k',\nu',0})}{p_y} \\ &\quad \le g_2(y)\exp\Big(-k \inf_{\theta \in \Theta_\varepsilon} \Big(H\big(\theta \mid (\theta)_1 \times \mu_X \times \mu_Y\big) - H\big((\theta)_2 \mid \mu_X \times \mu_Y\big)\Big)\Big), \end{aligned}$$

where $\Theta_\varepsilon = \{\theta \in M_1(\Sigma^2): (\theta)_1, (\theta)_2, (\theta)_{12} \in G_\varepsilon\}$ and $y^{-1}\log g_2(y) \to 0$ with y, independently of $e \in \mathcal{E}_c$ and of η.

It is easy to check that for all $\theta \in M_1(\Sigma^2)$,

$$H\big(\theta \mid (\theta)_1 \times \mu_X \times \mu_Y\big) - H\big((\theta)_2 \mid \mu_X \times \mu_Y\big) = H\big(\theta \mid (\theta)_1 \times (\theta)_2\big) \ge 0 \tag{2.25}$$

with equality iff $\theta = (\theta)_1 \times (\theta)_2$. Equality cannot be achieved in (2.25) when $(\theta)_1 = (\theta)_2 = (\theta)_{12} = \alpha^*$ because by (E′), $(\alpha^* \times \alpha^*)_{12} = \alpha_X^* \times \alpha_Y^* \ne \alpha^*$. In view of the continuity of $\theta \mapsto H(\theta \mid (\theta)_1 \times (\theta)_2)$ and the compactness of $M_1(\Sigma^2)$, it follows that for all $\varepsilon = \eta + 12\delta\|F\|_\infty$ small enough,

$$\beta'(\varepsilon) = \inf_{\theta \in \Theta_\varepsilon} \Big\{H\big(\theta \mid (\theta)_1 \times \mu_X \times \mu_Y\big) - H\big((\theta)_2 \mid \mu_X \times \mu_Y\big)\Big\} > 0.$$

This in turn implies, for η, δ small enough (again, the choice $\delta = \sqrt{\eta}$ with η small enough will do) and $\beta = \beta'(\varepsilon)/\|F\|_\infty > 0$, that for each $e \in \mathcal{E}_c$,

$$\frac{P(J_{x,k,\nu,\xi}, J_{x',k',\nu',0})}{p_y} \le g_2(y)\exp(-\beta y). \tag{2.26}$$

Combining now (2.18), (2.22) and (2.26), one sees that $\lim_{y\to\infty} Q_2(y) = 0$ [see (2.17)], completing the proof of the lemma. □

3. Proof of (1.8) and an example satisfying (E_Y).

PROOF OF (1.8). By (E_Y) and the continuity of $H(\cdot\,|\mu_X \times \mu_Y)$ there exists a relatively open subset U of $\{\nu\colon \nu(\overline{\Sigma}) = 1\}$ such that $U_Y = \{\nu_Y\colon \nu \in U\}$ is an open neighborhood of β^* and

$$\sup_{\nu \in U}\Big\{H\big(\nu \mid \mu_X \times \nu_Y\big) - H\big(\nu_Y \mid \mu_Y\big)\Big\} \le \frac{1-\delta}{1+\delta}H\big(\beta^* \mid \mu_Y\big),$$

for some $\delta > 0$. Let $I_n = \{\Delta\colon |H(\beta^* \mid \mu_Y)\Delta/\log n - 1| \le \delta\}$ and set $\Delta_n, j_n \le n - \Delta_n$ to be such that $\overline{R}_n^Y = \sum_{k=1}^{\Delta_n} G_Y(Y_{j_n+k})$. Note that $M_n = \overline{R}_n^Y$ if for some $i = 0, \dots, [n/\Delta_n]-1$ the empirical measure $L_{\Delta_n}^{T^{i\Delta_n}\mathbf{X},\, T^{j_n}\mathbf{Y}}$ of the pairings $(X_{i\Delta_n+k}, Y_{j_n+k})$ is supported on $\overline{\Sigma}$. By [8], Theorem 2,

$$q_n = P\Big(\Delta_n \in I_n, L_{\Delta_n}^{T^{j_n}\mathbf{Y}} \in U_Y\Big) \to_{n\to\infty} 1.$$

For n large enough, every $\Delta_n \in I_n$ and all i,

$$\begin{aligned} &P\Big(L_{\Delta_n}^{T^{i\Delta_n}\mathbf{X},\, T^{j_n}\mathbf{Y}} \in U \mid \Delta_{n,j_n}, L_{\Delta_n}^{T^{j_n}\mathbf{Y}} \in U_Y\Big)\\ &\qquad \ge (\Delta_n + 1)^{-(|\Sigma|-1)} \exp\Big(-\Delta_n(1-\delta)H\big(\beta^* \mid \mu_Y\big)/(1+\delta)\Big) = p(\Delta_n)\end{aligned}$$

(see [8], (3) and (5)). For some $c > 0$ and all n large enough, $\inf_{\Delta \in I_n}[n/\Delta]p(\Delta) \ge cn^{\delta/2}$. Hence, by the independence of $(X_{i\Delta_n+1}, \dots, X_{i\Delta_n+\Delta_n})$,

$$P\big(M_n = \overline{R}_n^Y\big) \ge q_n \inf_{\Delta \in I_n}\Big\{1 - \big(1 - p(\Delta)\big)^{[n/\Delta]}\Big\} \ge q_n\big(1 - e^{-cn^{\delta/2}}\big) \to_{n\to\infty} 1. \quad \square$$

The following example satisfies (E_Y) for $\Sigma_X = \Sigma_Y = \{0,1,2\}$. Let $\mu_X(i) = 1/3$, $i = 0,1,2$, $\mu_Y(0) = \mu_Y(1) = 1/6$ and consider the symmetric score $F(x,y) = 1$ for $x + y < 2$ while $F(x,y) = -\infty$ otherwise [so $F(x,y) \ne F(x) + F(y)$]. Here, $E_{\mu_Y}(G_Y) = -\infty$ and $\overline{\Sigma} = \{(0,0),(0,1),(1,0)\}$, with $\overline{\theta}^* = H(\beta^* \mid \mu_Y) = \log 3$, $\beta^*(0) = \beta^*(1) = 1/2$ and $E_\nu(F) = 1$ as soon as $\nu(\overline{\Sigma}) = 1$. Thus, (E_Y) holds since $H(\nu \mid \mu_X \times \mu_Y) < 2\log 3$ for $\nu((0,1)) = 1/2, \nu((0,0)) = \nu((1,0)) = 1/4$. In this particular example, $\theta^* = \log 6$, hence $1/\theta^* < \gamma^* < 2/\theta^*$, while $\mathcal{M} = \{\nu\colon \nu((0,1)) = 1/2, \nu((0,0)) + \nu((1,0)) = 1/2\}$ is the set of limit points of the empirical measures of pairings $(X_{i+\ell}, Y_{j+\ell})$ over the segment where M_n is achieved (cf. [8], Theorem 2). In particular, $|\mathcal{M}| = \infty, \alpha^* \notin \mathcal{M}$ and (E′) fails while M_n possesses a limit distribution of type I [up to lattice effetcs as in (1.4)]. $\square$

REFERENCES

[1] ARRATIA, R. and WATERMAN, M. S. (1994). A phase transition for the score in matching random sequences allowing deletions. *Ann. Appl. Probab.* **4** 200–225.

[2] ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

[3] ARRATIA, R., GORDON, L. and WATERMAN, M. S. (1986). An extreme value theory for sequence matching. *Ann. Statist.* **14** 971–993.

[4] ARRATIA, R., GORDON, L. and WATERMAN, M. S. (1990). The Erdös–Rényi law in distribution for coin tossing and sequence matching. *Ann. Statist.* **18** 539–570.
[5] CSISZÁR, I., COVER, T. M. and CHOI, B. S. (1987). Conditional limit theorems under Markov conditioning. *IEEE Trans. Inform. Theory* **33** 788–801.
[6] DEMBO, A. and KARLIN, S. (1991). Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables. *Ann. Probab.* **19** 1737–1755.
[7] DEMBO, A. and KARLIN, S. (1992). Poisson approximations for r-scan processes. *Ann. Appl. Probab.* **2** 329–357.
[8] DEMBO, A. KARLIN, S. and ZEITOUNI, O. (1994). Critical phenomena for sequence matching with scoring. *Ann. Probab.* **22** 1993–2021.
[9] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.
[10] IGLEHART, D. (1972). Extreme values in the $GI/G/1$ queue. *Ann. Math. Statist.* **43** 627–635.
[11] KARLIN, S. and DEMBO, A. (1992). Limit distributions of maximal segmental score among Markov dependent partial sums. *Adv. in Appl. Probab.* **24** 113–140.
[12] KARLIN, S. and OST, F. (1988). Maximal length of common words among random letter sequences. *Ann. Probab.* **16** 535–563.
[13] NEUHAUSER, C. (1994). A Poisson approximation for sequence comparisons with insertions and deletions. *Ann. Statist.* **22** 1603–1629.

A. DEMBO
DEPARTMENT OF MATHEMATICS
AND DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

S. KARLIN
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

O. ZEITOUNI
DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION–ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA 32000
ISRAEL