# Critical Phenomena for sequence matching with scoring

Amir Dembo*
Department of Mathematics and Department of Statistics
Stanford University
Stanford, CA 94305, U.S.A.

Samuel Karlin†
Department of Mathematics
Stanford University
Stanford, CA 94305 U.S.A.

Ofer Zeitouni††
Department of Electrical Engineering
Technion
Haifa 32000, Israel

4 November 1992; Revised: 7 January 1994

**Abstract**

Consider two independent sequences $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$. Suppose that $X_1, \cdots, X_n$ are i.i.d. $\mu_X$ and $Y_1, \cdots, Y_n$ are i.i.d. $\mu_Y$, where $\mu_X$ and $\mu_Y$ are distributions on finite alphabets $\Sigma_X$ and $\Sigma_Y$, respectively. A score $F : \Sigma_X \times \Sigma_Y \to I\!R$ is assigned to each pair $(X_i, Y_j)$ and the maximal non-aligned segment score is $M_n = \max_{\substack{0 \le i,j \le n-\Delta \\ \Delta \ge 0}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell})\}$. Our result is that $M_n / \log n \to \gamma^*(\mu_X, \mu_Y)$ a.s. with $\gamma^*$ determined by a tractable variational formula. Moreover, the pair empirical measure of $(X_{i+\ell}, Y_{j+\ell})$ during the segment where $M_n$ is achieved, converges to a probability measure $\nu^*$ which is accessible by the same formula. These results generalize to $X_i, Y_j$ taking values in any Polish space, to intrasequence scores under shifts, to long quality segments, and to more than two sequences.

## 1. Introduction.

Our motivation derives from biomolecular sequence comparisons. In DNA and protein sequence matching, segment scores are of the form $\sum_{\ell=1}^{\Delta} F(a_{i+\ell}, a'_{j+\ell})$, where $a_i$ is the $i$-th letter in the first sequence, $a'_j$ is the $j$-th letter in the second sequence and $F(x, y)$ is the score for the letter pair $(x, y)$. For the longest perfect match and the longest quality $q$ match (% matching exceeding $q$), the formulas of Karlin and Ost (1988), Arratia et al. (1986, 1990) determine the asymptotic distribution of the maximal intersequence segment score for matching with few errors, at least when the laws of the sequences are similar enough.

---

* Partially supported by NSF DMS92-09712 grant and by a US-ISRAEL BSF grant.

† Partially supported by grants NIH 5R01GM10452-29, 8R01HG00335-04 and NSF DMS86-06244.

‡ Partially supported by a US-ISRAEL BSF grant and Fund for promotion of research at the Technion.

**Key Words:** Large deviations, strong laws, sequence matching, large segmental sums.

AMS (1991) subject classifications: Primary: 60F10; Secondary: 60F15.

This paper is inspired by the critical phenomena in sequence matching introduced in Arratia and Waterman (1985), Arratia et al. (1988). It is important to seek a generalization of this phenomena to the context of general scoring systems. Thus, consider two sequences of length $n$, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, where the letters $X_i$ take values in a finite alphabet $\Sigma_X$ and the letters $Y_i$ take values in a finite alphabet $\Sigma_Y$. 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

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

Suppose the two sequences are independent: $X_1, \ldots, X_n$ i.i.d. following the distribution law $\mu_X$ and $Y_1, \ldots, Y_n$ i.i.d. $\mu_Y$, where $\mu_X$ and $\mu_Y$ refer to probabilities on $\Sigma_X$ and $\Sigma_Y$, 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. So we assume

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

in which case $M_n \to \infty$ corresponds to rare events.

The hypothesis (H) is in force throughout this paper, and it is also assumed that $\mu_X$ and $\mu_Y$ are strictly positive on $\Sigma_X$ and $\Sigma_Y$, respectively. The growth asymptotics of $M_n$ under (H) is characterized in the following theorem.

**Theorem 1.** *There exists a finite and positive constant $\gamma^*(\mu_X, \mu_Y)$ depending on $\mu_X$ and $\mu_Y$ such that*

$$M_n / \log n \to \gamma^*(\mu_X, \mu_Y) \ \text{ a.s.}$$

(throughout this paper log means the natural logarithm).

The constant $\gamma^*$ is expressed in terms of relative entropy functionals detailed in (1) below.

Vital for applications is the precise limit distribution of $M_n$ centered at $\gamma^* \log n$. This would provide the means for rigorous assessment of statistical significance in sequence comparisons with general scoring schemes of paramount relevance to the BLAST programs (e.g., Altschul et al. (1990), States et al. (1992)) and for algorithms founded on information (likelihood ratio) scoring matrices as in Stormo and Hartzell (1989) or Henikoff and Henikoff (1992). In the case of either perfect matching of similar sequences or maximal segmental score for aligned sequences, the limit distribution is an extremal distribution of type I (cf. Arratia et al. (1986), Karlin and Ost (1988), Karlin and Dembo (1992)). Under conditions related to (E) in Theorem 4 below and relying on the results of this paper, we prove in a companion paper (Dembo et al. (1994)) that the limit distribution is again an extremal distribution of type I. The relevant mean displacement constant is then tractable based on sums of i.i.d. fluctuation theory identities.

The following notations are used. Let $\Sigma = \Sigma_X \times \Sigma_Y$ be the (finite) alphabet of letter pairs, and let $M_1(\Sigma)$ denote the set of all probability measures on $\Sigma$ and $M_F(\Sigma)$ the subset of probability measures $\nu$ satisfying $E_\nu(F) \geq 0$. The relative entropy of $\nu \in M_1(\Sigma)$ with respect to $\mu \in M_1(\Sigma)$ is denoted by $H(\nu|\mu)$, and for $\Sigma = \{b_1, \cdots, b_N\}$ given by the formula:

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

with $0 \log 0$ interpreted as $0$.

The function $J : M_1(\Sigma) \to [-\infty, \infty)$ is defined via

$$J(\nu) = \frac{E_\nu(F)}{H^*(\nu|\mu_X, \mu_Y)}$$

where

$$H^*(\nu|\mu_X, \mu_Y) = \max\{\frac{1}{2}H(\nu|\mu_X \times \mu_Y), H(\nu_X|\mu_X), H(\nu_Y|\mu_Y)\}$$

and $\nu_X$ and $\nu_Y$ denote the marginals of $\nu$ on $\Sigma_X$ and $\Sigma_Y$, respectively. We shall write $H^*(\nu)$ for $H^*(\nu|\mu_X, \mu_Y)$ where no ambiguity concerning the measures $\mu_X$, $\mu_Y$ is likely. Note that $J(\nu)$ is finite except for $\nu = \mu_X \times \mu_Y$ in which case owing to (H) $J(\nu) = -\infty$. The constant $\gamma^*(\mu_X, \mu_Y)$ is proved (Theorem 3) to be

$$\gamma^*(\mu_X, \mu_Y) \equiv \sup_{\nu \in M_1(\Sigma)} J(\nu) = \sup_{\nu \in M_F(\Sigma)} J(\nu), \tag{1}$$

where the last equality is a consequence of (H). With $\mu_X \times \mu_Y$ strictly positive and $\Sigma$ finite, it follows that $H^*(\nu|\mu_X \times \mu_Y)$ is continuous with respect to $\nu$, and by (H) so is the extended real valued function $J(\nu)$. In particular, $\gamma^*(\mu_X, \mu_Y)$ is finite, the sets $\{\nu : J(\nu) \geq \beta\}$ are compact for all $\beta > 0$, and the compact set

$$\mathcal{M} = \{\nu : J(\nu) = \gamma^*(\mu_X, \mu_Y)\}$$

is non-empty. Let $\Delta_n^*$ denote the length of any segment where $M_n$ is achieved and let $\nu_n^*$ be the empirical measure of pairings $(X_{i+\ell}, Y_{j+\ell})$ over this segment.

The asymptotic properties of $\Delta_n^*$ and $\nu_n^*$ are described in the following theorem.

**Theorem 2.** *All limit points of $\nu_n^*$ belong to the set $\mathcal{M}$ a.s., and all limit points of $\Delta_n^*/\log n$ belong to the set $\{1/H^*(\nu) : \nu \in \mathcal{M}\}$.*

Both Theorems 1 and 2 are corollaries of the following more general result about the growth asymptotics of the maximal score among all segments of specified empirical measure of the pairs $(X_{i+\ell}, Y_{j+\ell})$.

**Theorem 3.** *For any $U \subset M_1(\Sigma)$ let $J_U = \max\{\sup_{\nu \in U} J(\nu), 0\}$ and*

$$M_n^U = \max\{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell}) : \Delta \geq 0,\ i, j \leq n - \Delta,\ L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})} \in U\}$$

*where* $T^i\mathbf{X} = (X_{i+1}, X_{i+2}, \cdots), T^j\mathbf{Y} = (Y_{j+1}, Y_{j+2}, \cdots)$ *and* $L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})}$ *is the empirical measure of the pairs* $\{(X_{i+\ell}, Y_{j+\ell})\}_{\ell=1}^{\Delta}$. *Then*

*(a)* $\limsup_{n\to\infty} M_n^U / \log n \leq J_U$ *a.s.*

*(b)* $\liminf_{n\to\infty} M_n^U / \log n \geq J_{U^o}$ *a.s. (where* $U^o$ *denotes the interior of the set* $U$*).*

Theorem 1 follows by applying Theorem 3 with $U = M_1(\Sigma)$, in which case $J_{M_1(\Sigma)} = \gamma^*$. Theorem 2 follows by observing that $M_n = \Delta_n^* E_{\nu_n^*}(F)$, and comparing Theorem 1 with the upper bound (a) of Theorem 3 applied to $U$ being the complement of the $\delta$-sphere blow-up of the set $\mathcal{M}$.

Examination of the proof of Theorem 3 reveals that the set $U^o$ in part (b) of Theorem 3 can actually be replaced by any subset of $V$ – the set of all limits of sequences $\{\nu_k\} \subset U$ for which the vectors $k\nu_k$ have integer coordinates for $k = 1, 2, \ldots$. Theorem 3 is thus useful in accommodating scoring schemes in which some pair letter combinations are forbidden, i.e. when $-\infty$ is in the range of $F$. For example, Theorem 1 is then deduced by applying Theorem 3 with $U = V = \{\nu : E_\nu(F) > -\infty\}$. The sequence matching problems discussed in Arratia and Waterman (1985) and in Arratia et al. (1988) are examples of such schemes.

Of particular interest are the conditions under which $\mathcal{M} = \{\nu^*\}$ is unique (for then Theorem 2 states that $\nu_n^* \to \nu^*$ a.s. and $\Delta_n^* / \log n \to 1/H^*(\nu^*)$ a.s.), as well as providing simpler characterizations of both $\gamma^*$ and $\nu^*$. Along these lines, recall that (H) entails the existence of a unique positive value $\theta^*$ satisfying

$$E_{\mu_X \times \mu_Y}\left[e^{\theta^* F}\right] = 1,$$

and let $\alpha^* \in M_1(\Sigma)$ denote the conjugate measure associated with $\theta^*$, i.e.

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

Henceforth, $\alpha_X^*$ and $\alpha_Y^*$ denote the marginals of $\alpha^*$ on $\Sigma_X$ and $\Sigma_Y$, respectively.

The maximal segment score without shifts, is

$$R_n = \max_{\substack{0 \leq i \leq n-\Delta \\ \Delta \geq 0}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{i+\ell})\}.$$

It is not hard to check that (H) implies that $R_n / \log n \to 1/\theta^*$ a.s., and as shown in Dembo and Karlin (1991a) the pair empirical measure of $(X_{i+\ell}, Y_{i+\ell})$ over the segment where $R_n$ is achieved, converges a.s. to $\alpha^*$. Intuitively, one might expect $M_n$ to be comparable to $R_{n^2}$ and concomitantly that $\gamma^* = 2/\theta^*$ and $\nu^* = \alpha^*$. While in general this is not the case, the following theorem states an accessible necessary and sufficient condition for $\gamma^* = 2/\theta^*$ and $\nu^* = \alpha^*$.

**Theorem 4.** *Always* $\gamma^*(\mu_X, \mu_Y) \leq 2/\theta^*$. *The following conditions are equivalent:*

*(a)* $\gamma^*(\mu_X, \mu_Y) = 2/\theta^*$ *and* $\mathcal{M} = \{\alpha^*\}$.

*(b)* $(\mu_X, \mu_Y)$ *satisfy the inequality*

$$\text{(E)} \qquad H(\alpha^*|\mu_X \times \mu_Y) \geq 2\max\{H(\alpha_X^*|\mu_X), H(\alpha_Y^*|\mu_Y)\}.$$

*In particular, for identical alphabets and symmetric scores* $(F(x,y) = F(y,x))$*, condition (E) holds whenever* $\mu_X = \mu_Y$*, and unless* $F(x,y)$ *is of the form* $F(x)+F(y)$*, condition (E) also applies to all* $\mu_X \times \mu_Y$ *in some open neighborhood of the diagonal* $\{\mu \times \mu : \mu \in M_1(\Sigma_X)\ ,\ \mu \times \mu$ *satisfying* $(H)$ $\}$.

**Remark.** If $F(x,y)$ is a function of $x$ only, then $M_n = R_n$ and hence $\gamma^*(\mu_X, \mu_Y) = 1/\theta^*$ for any $\mu_X$, $\mu_Y$ satisfying (H). This shows that some symmetry condition on $F(x,y)$ is needed for (E) to hold. Another example concerns $\Sigma_X = \Sigma_Y = \{-1, 1\}$ with $F(x,y) = F(x) + F(y) = x + y$. Then, (H) is satisfied as soon as $0 < \mu_X(1) < 1/2$ and $0 < \mu_Y(1) < 1/2$. On the other hand, since $\alpha^*$ is a product measure, it follows that $H(\alpha^*|\mu_X \times \mu_Y) = H(\alpha_X^*|\mu_X) + H(\alpha_Y^*|\mu_Y)$, and (E) holds iff $H(\alpha_X^*|\mu_X) = H(\alpha_Y^*|\mu_Y)$. Letting $p = \mu_X(1)$ and $\pi = \mu_Y(1)$, one finds that

$$\theta^* = \frac{1}{2}\log(\frac{(1-p)(1-\pi)}{p\pi}).$$

With $p$ fixed, define $f(\pi) = H(\alpha_X^*|\mu_X) - H(\alpha_Y^*|\mu_Y)$. One finds directly that $f(p) = 0$ and when $p < 1/2$, we have $f''(y) < 0$ for all $y < 1/2$ and $f(1/2) > 0$. Thus, $f(\pi) = 0$ iff $\pi = p$, i.e. (E) holds iff $\mu_X = \mu_Y$.

Even when (E) fails, uniqueness of the measure $\nu^*$ yielding the value $\gamma^*$ is assured in the following case.

**Theorem 5.** *Let* $G_X(x) = \max_y\{F(x,y)\}$, $\quad G_Y(y) = \max_x\{F(x,y)\}$. *Then,* $\mathcal{M} = \{\nu^*\}$ *when the following conditions hold:*

*(a) Either* $E_{\mu_X}(G_X) \geq 0$*, or* $G_X(x) = F(x,y)$ *has a unique solution* $y(x)$ *for all* $x$*; and,*

*(b) Either* $E_{\mu_Y}(G_Y) \geq 0$*, or* $G_Y(y) = F(x,y)$ *has a unique solution* $x(y)$ *for all* $y$*.*

**Remarks:** (a) Suppose $F(x,y) = F(x)$ is a function of $x$ only. Then, $\alpha_Y^* = \mu_Y$, and it is not hard to check that $\{\mathcal{M}\}$ contains all measures of the form $\nu = \alpha_X^* \times \nu_Y$ with $\nu_Y \in M_1(\Sigma_Y)$ such that $H(\nu_Y|\mu_Y) \leq H(\alpha_X^*|\mu_X)$. In this example $\mathcal{M}$ contains infinitely many probability measures, while $G_X(x) = F(x)$ and because of (H) condition (a) of Theorem 5 is violated.

(b) The conditions of Theorem 5 are trivially met in matching problems when $F(x,x) > 0$ and $F(x,y) < 0$ for $x \neq y$.

(c) Let $S$ be any closed set of strictly positive measures $\mu_X \times \mu_Y$ satisfying (H). Since $\gamma^*(\mu_X, \mu_Y)$ is positive throughout the compact set $S$, it is not hard to check that there exists $\delta = \delta(S) > 0$ such that for all $\mu_X \times \mu_Y \in S$,

$$\gamma^*(\mu_X, \mu_Y) = \sup_{\{\nu : H(\nu|\mu_X \times \mu_Y) \geq \delta\}} J(\nu) \quad .$$

With $\Sigma$ finite, and $F(x,y)$ bounded, the function $J(\nu)$ is (uniformly) continuous jointly in $\nu$ and $\mu_X \times \mu_Y$, on the set $\{(\nu, \mu_X \times \mu_Y) : H(\nu|\mu_X \times \mu_Y) \geq \delta,\ \mu_X \times \mu_Y \in S\}$. Therefore, $\gamma^*(\mu_X, \mu_Y)$ is continuous on $S$, and if $\mathcal{M} = \{\nu^*\}$ for every $\mu_X \times \mu_Y \in S$, then $\nu^* = \nu^*(\mu_X, \mu_Y)$ is also continuous on $S$.

The next two sections are devoted to the proofs of Theorems 3, 4 and 5. They are followed by several extensions. In biomolecular sequence analysis the two (or more) sequences often have very different lengths. Sequences of possibly different lengths are the focus of Section 4, while scores for more than two independent sequences are considered in Section 5. Theorems 1–5 are shown in Section 6 to apply for intrasequence scores with shifts, i.e.

$$M_n = \max_{\substack{0 \leq i \neq j \leq n-\Delta \\ \Delta \geq 0}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, X_{j+\ell})\},$$

encompassing the context of significant repeat segments. The asymptotics of the longest quality $q$ match is considered in Arratia and Waterman (1989) and Arratia et al. (1990). In Section 7, we present this problem and its extension to a vector-valued scoring system, and relate them with the results of Sections 3.2 and 5.5 of Dembo and Zeitouni (1993). Section 8 concerns extensions of Theorems 1,2,4 and 5 to infinite alphabets. Under appropriate modifications, such extensions apply also to the formulations of Sections 4-7. By fixing the alphabet of the second sequence to be a singleton, the theory is specialized to a one sequence setting, in particular providing an alternative simpler proof of the strong laws of Dembo and Karlin (1991a).

When the i.i.d. assumption is generalized to Markov dependence, the corresponding results for long segments of perfect matching or of quality $q$ are reported in Arratia and Waterman (1985,1989), in Karlin and Ost (1988), and in Arratia et al. (1988,1990), while for general scoring systems with aligned sequences such results are given in Dembo and Karlin (1991b).

## 2. Proof of Theorem 3.

The proof of the theorem starts with the following simple lemma which allows us to restrict attention thereafter to segments of length at most $O(\log n)$.

*Lemma 1.* There exists $c_0$ large enough, such that for all $n$,

$$P\left[\sup_{\substack{0 \leq i,j \leq n-\Delta \\ \Delta \geq c_0 \log n}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell})\} \geq 0\right] \leq 1/n^2 \ .$$

Proof: Since there are at most $n^3$ possible choices of $i, j, \Delta$, by the union of events bound (and stationarity of the processes $\{X_\ell\}$ and $\{Y_\ell\}$), the lemma will follow from the tail estimate

$$\sup_{\Delta \geq c_0 \log n} P(\sum_{\ell=1}^{\Delta} F(X_\ell, Y_\ell) \geq 0) \leq 1/n^5 \ . \tag{2}$$

Let

$$\Lambda(\lambda) \equiv \log E_{\mu_X \times \mu_Y}[e^{\lambda F}].$$

Then, for $\lambda \geq 0$, by Chebychev's bound

$$P(\sum_{\ell=1}^{\Delta} F(X_\ell, Y_\ell) \geq 0) \leq E[\exp(\lambda \sum_{\ell=1}^{\Delta} F(X_\ell, Y_\ell))] = (E[e^{\lambda F}])^{\Delta} \leq e^{\Delta \Lambda(\lambda)} \ .$$

From (H), $E_{\mu_X \times \mu_Y}(F) = \Lambda'(0) < 0$, and therefore there exists $\lambda_0 > 0$ (in fact for all $0 < \lambda_0 < \theta^*$), such that $-\Lambda(\lambda_0) > 0$. By choosing $\lambda = \lambda_0$ and $c_0 = 5/(-\Lambda(\lambda_0))$, (2) ensues. ∎

While Lemma 1 applies to any alphabet $\Sigma$, we shall hereafter exploit finiteness of $\Sigma$, and predicate the proof of the theorem on the method of types. For a full account of this method and its applications, see Csiszár and Körner (1981) Chapter 1, Cover and Thomas (1991) Chapter 12, or Dembo and Zeitouni (1993) Section 2.1. We recall here only the results needed. Let $\mathcal{L}_k$ be the set of probability vectors $\nu$ on $\Sigma$ for which all the coordinates of $k\nu$ are integers. With $L_k^{(X,Y)}$ denoting the empirical measures of pairs $\{(X_\ell, Y_\ell)\}_{\ell=1}^k$, the following estimates are easy consequences of the attendant multinomial probabilities:

$$(3) \qquad (k+1)^{-(|\Sigma|-1)} e^{-kH(\nu|\mu_X \times \mu_Y)} \leq P(L_k^{(X,Y)} = \nu) \leq e^{-kH(\nu|\mu_X \times \mu_Y)} \quad \forall \nu \in \mathcal{L}_k,$$

with $|\Sigma|$ denoting the cardinality of the set $\Sigma$.

Similar estimates hold for $L_k^X$ and $L_k^Y$, the empirical measures of $\{X_\ell\}_{\ell=1}^k$ and of $\{Y_\ell\}_{\ell=1}^k$, respectively. Specifically,

$$(4) \qquad (k+1)^{-(|\Sigma|-1)} e^{-kH(\nu_X|\mu_X)} \leq P(L_k^X = \nu_X) \leq e^{-kH(\nu_X|\mu_X)}$$

and

$$(5) \qquad (k+1)^{-(|\Sigma|-1)} e^{-kH(\nu_Y|\mu_Y)} \leq P(L_k^Y = \nu_Y) \leq e^{-kH(\nu_Y|\mu_Y)}.$$

Finally,

$$(6) \qquad |\mathcal{L}_k| \leq (k+1)^{|\Sigma|-1},$$

and for all $\nu \in M_1(\Sigma)$

$$(7) \qquad \inf_{\tilde{\nu} \in \mathcal{L}_k} \|\nu - \tilde{\nu}\|_{\text{var}} \leq |\Sigma|/k$$

(with $m_i/k \leq \nu(a_i) < m_i/k + 1/k$, we have $\sum_i m_i \leq k$. Let $\tilde{\nu}(a_i) = m_i/k$, adding at most $1/k$ to each coordinate till $\sum_i \tilde{\nu}(a_i) = 1$).

We turn now to prove the upper bound on $M_n^U$.

_Lemma 2._ Let $\bar{M}_n^U$ denote the value of $M_n^U$ confined to segments of length at most $c_0 \log n$ (with $c_0$ as determined by Lemma 1). Suppose that $J_U > 0$. Then, for all $t > 1$

$$P(\bar{M}_n^U \geq t J_U \log n) \leq \frac{(c_0 \log n + 1)^{|\Sigma|}}{n^{t-1}}.$$

Proof: For $\nu \in \mathcal{L}_k$, $k \le c_0 \log n$ we consider the events

(8) $$A_{\nu,k} = \{\exists i,j : \quad 0 \le i,j \le n-k \text{ such that } L_k^{(T^i X, T^j Y)} = \nu\} \ .$$

($A_{\nu,k}$ is the event that there exists a segment of length $k$ in which the empirical measure of pairs $\{(X_{i+\ell}, Y_{j+\ell})\}_{\ell=1}^k$ is $\nu$). Applying the union of events bound with respect to the values of $0 \le i,j \le n-1$,

$$P(A_{\nu,k}) \le \min\{n^2 P(L_k^{(X,Y)} = \nu), nP(L_k^X = \nu_X), nP(L_k^Y = \nu_Y), 1\},$$

and by the upper estimates of (3)–(5),

$$P(A_{\nu,k}) \le \min\{n^2 e^{-kH(\nu|\mu_X \times \mu_Y)}, ne^{-kH(\nu_X|\mu_X)}, ne^{-kH(\nu_Y|\mu_Y)}, 1\}.$$

Consequently, using the inequality $\min\{a^2, 1\} \le a$ for $a > 0$, and the definition $H^*(\nu) = \max\{\frac{1}{2}H(\nu|\mu_X \times \mu_Y), H(\nu_X|\mu_X), H(\nu_Y|\mu_Y)\}$, we have

$$P(A_{\nu,k}) \le ne^{-kH^*(\nu)} \ .$$

Hence,

(9) $$kH^*(\nu) \ge t \log n \quad \Rightarrow \quad P(A_{\nu,k}) \le n^{-(t-1)} \ .$$

We partition the event $\{\bar{M}_n^U \ge tJ_U \log n\}$ according to the pair empirical measure $\nu$ on the segment in which the score $\bar{M}_n^U$ is obtained, observing that the score is $\sum_{\ell=1}^k F(X_{i+\ell}, Y_{j+\ell}) = kE_\nu(F)$. Thus,

$$\{\bar{M}_n^U \ge tJ_U \log n\} = \bigcup_{\left\{\substack{k \le c_0 \log n \\ \nu \in \mathcal{L}_k, \nu \in U \\ kE_\nu(F) \ge tJ_U \log n}\right\}} A_{\nu,k} \ .$$

Since $J_U > 0$, $kE_\nu(F) \ge tJ_U \log n$ for $\nu \in U$ implies that $kH^*(\nu) \ge t \log n$. Thus, with $\sum_{k \le c_0 \log n} |\mathcal{L}_k| \le (c_0 \log n + 1)^{|\Sigma|}$, the proof of the lemma is completed by applying the union of events bound in conjunction with the estimate of (9). ∎

Turning now to the lower bound on $M_n^U$, we first prove the following estimate on the basic events $A_{\nu,k}$ (see (8)).

*Lemma 3.* For any $\nu \in \mathcal{L}_k$, and any $n \ge k$,

(10)
$$1 - P(A_{\nu,k}) \le 4(k+1)^{(|\Sigma|+1)} n_*^{-2} e^{kH(\nu|\mu_X \times \mu_Y)} + (k+1)^{|\Sigma|} n_*^{-1} e^{kH(\nu_X|\mu_X)} + (k+1)^{|\Sigma|} n_*^{-1} e^{kH(\nu_Y|\mu_Y)} \ ,$$

where $n_* = k\lfloor \frac{n}{k} \rfloor$ is the largest integer multiple of $k$ not exceeding $n$.

Proof: Since $P(A_{\nu,k})$ is monotone increasing with $n$, without loss of generality, we may assume throughout that $n = n_*$ is an integer multiple of $k$. Divide the sequence $X_1, \cdots, X_n$ into $\frac{n}{k}$ disjoint successive segments of size $k$ each and let $N_X$ count the number of segments with empirical measure $\nu_X$. Similarly, let $N_Y$ count the number of corresponding segments from the sequence $Y_1, \cdots, Y_n$ of empirical measure $\nu_Y$. Conditioned on $N_X$ and $N_Y$ let $B_{ij}$ be the event that the $i$-th segment in the $X$ sequence with empirical measure $\nu_X$ and the $j$-th segment in the $Y$ sequence with empirical measure $\nu_Y$ are arranged in such a way that their corresponding letter pairs $(X, Y)$ empirical measure is precisely $\nu$.

Clearly, $B_{ij}$, $i = 1, \cdots, N_X, j = 1, \cdots, N_Y$ are events of equal probability. Therefore, with $W = \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} 1_{B_{ij}}$ and $p = P(B_{ij})$ we have

$$E[W|N_X, N_Y] = pN_X N_Y \ .$$

Now, given a segment in the $X$ sequence of empirical measure $\nu_X$ and a segment in the $Y$ sequence of empirical measure $\nu_Y$, the probability of associated letter pairs having empirical measure $\nu$ is independent of the precise order of the $X$-segment letters. Consequently, conditioned on $N_X$ and $N_Y$ the events $\{B_{ij}\}$ are pairwise independent, and in particular,

$$\mathrm{VAR}[W|N_X, N_Y] = N_X N_Y \mathrm{VAR}(1_{B_{ij}}) \leq pN_X N_Y.$$

Thus, by a Chebychev type bound

$$P[W = 0|N_X, N_Y] \leq E[\frac{(W - E[W|N_X, N_Y])^2}{E[W|N_X, N_Y]^2}|N_X, N_Y] \leq \frac{1}{pN_X N_Y}.$$

Hence, also

$$\begin{aligned} P[W = 0] &\leq E[P[W = 0|N_X, N_Y]1_{N_X \geq 1}1_{N_Y \geq 1}] + P(N_X = 0) + P(N_Y = 0) \\ &\leq \frac{1}{p} E[\frac{1}{N_X N_Y} 1_{N_X \geq 1} 1_{N_Y \geq 1}] + P(N_X = 0) + P(N_Y = 0) \\ &= \frac{1}{p} E[\frac{1_{N_X \geq 1}}{N_X}] E[\frac{1_{N_Y \geq 1}}{N_Y}] + P(N_X = 0) + P(N_Y = 0) \end{aligned}$$

where the equality is due to the independence of $N_X$ and $N_Y$. By definition $N_X \sim \text{Binomial}(\frac{n}{k}, p_X)$ and $N_Y \sim \text{Binomial}(\frac{n}{k}, p_Y)$ with $p_X = P(L_k^X = \nu_X)$ and $p_Y = P(L_k^Y = \nu_Y)$. Hence, by direct computation,

$$E[\frac{1_{N_X \geq 1}}{N_X}] \leq E[\frac{2}{N_X + 1}] = \frac{2(1 - (1 - p_X)^{n/k+1})}{p_X(\frac{n}{k} + 1)} \leq \frac{2k}{np_X},$$

with the analogous bound for $E[\frac{1_{N_Y \geq 1}}{N_Y}]$. Therefore, since $ay(1 - a)^y \leq 1$ for all $y > 0$, $0 \leq a \leq 1$,

$$P(W = 0) \leq \frac{4k^2}{n^2 pp_X p_Y} + (1 - p_X)^{\frac{n}{k}} + (1 - p_Y)^{\frac{n}{k}} \leq \frac{4k^2}{n^2 pp_X p_Y} + \frac{k}{np_X} + \frac{k}{np_Y}$$

Since $\{W \geq 1\} \subset A_{\nu,k}$, we have $1 - P(A_{\nu,k}) \leq P(W = 0)$. Consequently, observing that $pp_X p_Y = P(L_k^{(X,Y)} = \nu)$, we get

$$1 - P(A_{\nu,k}) \leq P(W = 0) \leq \frac{4k^2}{n^2 P(L_k^{(X,Y)} = \nu)} + \frac{k}{nP(L_k^X = \nu_X)} + \frac{k}{nP(L_k^Y = \nu_Y)},$$

and (10) follows by utilizing the lower bounds of (3)-(5). ∎

The estimates of Lemma 3 are combined next to achieve a lower bound on $M_n^U$.

*Lemma 4*. Suppose that $J_{U^o} > 0$. Then, for each $t < 1$, there exists an $n_0(t)$ such that for $n \geq n_0(t)$,

$$P(M_n^U \leq tJ_{U^o} \log n) \leq \frac{1}{n^{(1-t)/2}}. \tag{11}$$

Proof: Fix $t < 1$. Set $\tau = (1+2t)/3 < 1$, and note that $(\tau + t)/(2\tau) < 1$. Determine $\tilde{\nu} \in U^o$ with $J(\tilde{\nu}) > (\frac{\tau+t}{2\tau})J_{U^o}$. Let $k_n = \lceil \tau \log n / H^*(\tilde{\nu}) \rceil$ and recall that there exist $\tilde{\nu}_n \in \mathcal{L}_{k_n}$ with $\|\tilde{\nu} - \tilde{\nu}_n\|_{\text{var}} \leq |\Sigma|/k_n$ (see (7)). Hence, with $F(\cdot,\cdot)$ bounded, it follows that for some $C < \infty$ independent of $n$,

$$k_n E_{\tilde{\nu}_n}(F) \geq \tau \frac{E_{\tilde{\nu}_n}(F)}{H^*(\tilde{\nu})} \log n \geq \tau \frac{E_{\tilde{\nu}}(F)}{H^*(\tilde{\nu})} \log n - C = \tau J(\tilde{\nu}) \log n - C \geq (\frac{\tau + t}{2}) J_{U^o} \log n - C.$$

Since $\tilde{\nu} \in U^o$, so is $\tilde{\nu}_n$ for $n$ large enough, and moreover

$$k_n E_{\tilde{\nu}_n}(F) > tJ_{U^o} \log n$$

(since $\tau > t$). Consequently, for $n$ large enough, the event $\{M_n^U \leq tJ_{U^o} \log n\}$ is contained in the complement of $A_{\tilde{\nu}_n, k_n}$.

Thus, in view of the bound of (10) and the definition of $H^*(\nu)$ it suffices to show that

$$I_n = n^{(1-t)/2} (k_n + 1)^{|\Sigma|} n^{-1} e^{k_n H^*(\tilde{\nu}_n)} \to 0 \quad \text{as} \quad n \to \infty \,.$$

As $n \to \infty$, $\tilde{\nu}_n \to \tilde{\nu}$, and consequently $H^*(\tilde{\nu}_n) \to H^*(\tilde{\nu})$. Therefore, the term $I_n$ is of order $n^{(\tau - (1+t)/2 + o(1))}$. The proof of the lemma is complete, since $\tau < (1+t)/2$. ∎

Proof of Theorem 3: (a) Assume first that $J_U > 0$. Then, combining Lemmas 1 and 2, it follows that for any $1 < t < 3$ and all $n \geq n_0(t)$,

$$P(M_n^U \geq tJ_U \log n) \leq \frac{1}{n^{(t-1)/2}} \,. \tag{12}$$

Fix $t = 1 + \epsilon$, $\epsilon > 0$. Applying the Borel-Cantelli lemma along the skeleton $n_k = e^k$, it follows that a.s., $M_{n_k}^U \leq (1+\epsilon) J_U k$ for all $k \geq k_0(\omega)$. Increase $k_0(\omega)$ such that $k_0(\omega) \geq (1/\epsilon + 1)$.

Since $M_n^U$ is monotone in $n$, for all $n \geq n_0(\omega) = e^{k_0(\omega)}$

$$M_n^U \leq (1+\epsilon) J_U \lceil \log n \rceil \leq (1 + 2\epsilon) J_U \log n,$$

and consequently

$$\limsup_{n \to \infty} M_n^U / \log n \leq (1+2\epsilon) J_U \text{ a.s.}$$

With $\epsilon > 0$ arbitrarily small, the proof is complete.

If $J_U = 0$, then for all $\nu \in U$, $E_\nu(F) \le 0$ and trivially $M_n^U \le 0$ for all n.

(b) There is nothing to prove when $J_{U^o} = 0$. For $J_{U^o} > 0$ the proof is based on Lemma 4, paraphrasing the argument of part (a) above while changing $\epsilon$ to $-\epsilon$, $J_U$ to $J_{U^o}$, reversing direction of all inequalities involving $M_n^U$ and replacing $\lceil \log n \rceil$ with $\lfloor \log n \rfloor$.

**3. Proof of Theorems 4 and 5.**

In the proofs of Theorems 4 and 5, we shall use the well known information inequality

$$H(\nu|\mu_X \times \mu_Y) \ge H(\nu_X|\mu_X) + H(\nu_Y|\mu_Y), \tag{13}$$

with equality iff $\nu = \nu_X \times \nu_Y$ (considering $H(\nu_X|\mu_X) + H(\nu_Y|\mu_Y) - H(\nu|\mu_X \times \mu_Y)$, (13) is an immediate consequence of the concavity of the function $\log x$).

<u>Proof of Theorem 4</u>: Recall that $\alpha^*$ is defined via $\frac{d\alpha^*}{d(\mu_X \times \mu_Y)} = e^{\theta^* F}$; hence for any $\nu \in M_1(\Sigma)$

$$H(\nu|\mu_X \times \mu_Y) - \theta^* E_\nu(F) = H(\nu|\alpha^*) \; .$$

Of course, $H(\nu|\alpha^*) \ge 0$ with equality iff $\nu = \alpha^*$, and consequently for all $\nu \ne \alpha^*$,

$$H(\nu|\mu_X \times \mu_Y) > \theta^* E_\nu(F),$$

implying that $J(\nu) < 2/\theta^*$. Moreover, $H(\alpha^*|\mu_X \times \mu_Y) = \theta^* E_{\alpha^*}(F)$, and thus condition (E) is necessary and sufficient for $J(\alpha^*) = 2/\theta^*$. Consequently, $\gamma^*(\mu_X, \mu_Y) = 2/\theta^*$ iff (E) holds in which case $\mathcal{M} = \{\alpha^*\}$, and otherwise $\gamma^*(\mu_X, \mu_Y) < 2/\theta^*$.

Suppose now that $\Sigma_X = \Sigma_Y$ and $F(x,y) = F(y,x)$. Then, when $\mu_X = \mu_Y$ it follows that $\alpha_X^* = \alpha_Y^*$ and in particular $H(\alpha_X^*|\mu_X) = H(\alpha_Y^*|\mu_Y)$. In this case applying (13) for $\nu = \alpha^*$ yields (E). Moreover, (E) holds with equality only when $\alpha^* = \alpha_X^* \times \alpha_Y^*$. If this is the case, then setting $F_X = \frac{1}{\theta^*} \log \frac{d\alpha_X^*}{d\mu_X}$ and $F_Y = \frac{1}{\theta^*} \log \frac{d\alpha_Y^*}{d\mu_Y}$, it follows that $F(x,y) = F_X(x) + F_Y(y)$, and furthermore, by the symmetry of $F$ necessarily $F_X = F_Y$.

Assume now that $F(x,y)$ is symmetric but $F(x,y) \ne F(x) + F(y)$, so that (E) holds with strict inequality for $\mu_X = \mu_Y$. Since $F$ is bounded, $\theta^*$ is a continuous functional of $\mu_X \times \mu_Y$ and letting

$$\Phi(\mu_X \times \mu_Y) = H(\alpha^*|\mu_X \times \mu_Y) - 2\max\{H(\alpha_X^*|\mu_X), H(\alpha_Y^*|\mu_Y)\},$$

it follows that $\Phi$ is also a continuous functional of $\mu_X \times \mu_Y$.

Since $\Phi(\mu \times \mu) > 0$, it follows in particular that $\Phi(\mu_X \times \mu_Y) \ge 0$ (i.e. (E) holds), for some open neighborhood of the diagonal $\{\mu \times \mu : \mu \in M_1(\Sigma_X),\, \mu \times \mu \text{ satisfying (H)}\}$. ∎

The proof of Theorem 5 is based on the following lemma.

<u>*Lemma 5*</u>. Let $J_X(\nu) = E_\nu(F)/H(\nu_X|\mu_X)$ and $G_X(x) = \max_y F(x,y)$.

(a) If $E_{\mu_X}(G_X) \geq 0$, then $\sup_{\nu \in M_1(\Sigma)} J_X(\nu) = \infty$.

(b) If $E_{\mu_X}(G_X) < 0$ and for every $x \in \Sigma_X$, $G_X(x) = F(x,y)$ at a unique $y(x)$, then $J_X(\nu)$ has a unique global maximum.

Proof: Observe first that for a fixed $\sigma \in M_1(\Sigma_X)$,

$$\sup_{\nu_X = \sigma} J_X(\nu) = E_\sigma(G_X)/H(\sigma|\mu_X) \ . \tag{14}$$

Proof of (a): When $E_{\mu_X}(G_X) > 0$, the right hand side of (14) is infinity at $\sigma = \mu_X$. When $E_{\mu_X}(G_X) = 0$, let $\sigma_\epsilon$ be given by

$$\frac{d\sigma_\epsilon}{d\mu_X} = 1 + \epsilon(\eta 1_{G_X \geq 0} - 1_{G_X < 0})$$

with $\eta = \mu_X(G_X < 0)/\mu_X(G_X \geq 0)$. Then, $E_{\sigma_\epsilon}(G_X)/H(\sigma_\epsilon|\mu_X) \to \infty$ as $\epsilon \to 0$.
Proof of (b): Let $\theta_X^*$ denote the unique positive solution of $E_{\mu_X}[e^{\theta G_X}] = 1$. Define $\alpha_X \in M_1(\Sigma_X)$ via $\frac{d\alpha_X}{d\mu_X} = e^{\theta_X^* G_X}$. Then, for any $\sigma \in M_1(\Sigma_X)$

$$H(\sigma|\mu_X) - \theta_X^* E_\sigma(G_X) = H(\sigma|\alpha_X) \ ,$$

where $H(\sigma|\alpha_X) \geq 0$ with equality iff $\sigma = \alpha_X$. Hence, the right hand side of (14) attains a unique global maximum at $\sigma = \alpha_X$. The existence of a unique $y(x)$ for which $G_X(x) = F(x, y(x))$ requires that the unique global maximum of $J_X(\cdot)$ is at the measure $\nu$ characterized by $\nu(A) = \alpha_X(\{x : (x, y(x)) \in A\})$ for every $A \subset \Sigma$. ∎

Let $J_Y(\nu) = E_\nu(F)/H(\nu_Y|\mu_Y)$ and $J_{XY}(\nu) = 2E_\nu(F)/H(\nu|\mu_X \times \mu_Y)$. Lemma 5 has a counterpart statement in terms of $J_Y(\cdot)$, $E_{\mu_Y}(G_Y)$ and the solutions $x(y)$ of $G_Y(y) = \max_x F(x,y) = F(x(y), y)$. Thus under the conditions of Theorem 5, $J_X(\cdot)$ and $J_Y(\cdot)$ are either unbounded or have, respectively, unique global maxima.

Proof of Theorem 5: For any distinct $\nu_1$ and $\nu_2$ with $E_{\nu_1}(F) > 0$ and $E_{\nu_2}(F) > 0$, let $\nu_\lambda \equiv \lambda\nu_1 + (1-\lambda)\nu_2$ for $0 < \lambda < 1$. The linearity of the map $\nu \mapsto E_\nu(F)$ and the strict convexity of $H(\cdot|\mu_X \times \mu_Y)$ lead to the quasi-concavity of $J_{XY}(\cdot)$, i.e.

$$J_{XY}(\nu_\lambda) > \frac{2\lambda E_{\nu_1}(F) + 2(1-\lambda)E_{\nu_2}(F)}{\lambda H(\nu_1|\mu_X \times \mu_Y) + (1-\lambda)H(\nu_2|\mu_X \times \mu_Y)} \geq \min\{J_{XY}(\nu_1), J_{XY}(\nu_2)\}. \tag{15}$$

Similarly, it follows that

$$J_X(\nu_\lambda) \geq \min\{J_X(\nu_1), J_X(\nu_2)\}, \tag{16}$$

with strict inequality provided $J_X(\nu_1) \neq J_X(\nu_2)$, and

$$J_Y(\nu_\lambda) \geq \min\{J_Y(\nu_1), J_Y(\nu_2)\}, \tag{17}$$

with strict inequality provided $J_Y(\nu_1) \neq J_Y(\nu_2)$. Now suppose that there exist distinct $\nu_1$ and $\nu_2$ in $\mathcal{M}$, that is

$$\gamma^*(\mu_X, \mu_Y) = J(\nu_i) = \min\{J_X(\nu_i), J_Y(\nu_i), J_{XY}(\nu_i)\} \quad i = 1, 2. \tag{18}$$

Recall that $\gamma^*(\mu_X, \mu_Y) > 0$, which implies that $E_{\nu_i}(F) > 0$, and hence also $E_{\nu_\lambda}(F) > 0$. Comparing (18) with (15) it follows that $J_{XY}(\nu_\lambda) > \gamma^*(\mu_X, \mu_Y)$. Of course, $J(\nu_\lambda) \leq \gamma^*(\mu_X, \mu_Y)$, implying that $J_{XY}(\nu_\lambda) > J(\nu_\lambda)$, for all $0 < \lambda < 1$. By (13), this requires either $J_X(\nu_\lambda) < J_{XY}(\nu_\lambda) < J_Y(\nu_\lambda)$ or $J_Y(\nu_\lambda) < J_{XY}(\nu_\lambda) < J_X(\nu_\lambda)$. Moreover, from (16) and (17) we must have either

(i) $\gamma^*(\mu_X, \mu_Y) = J_X(\nu_\lambda) < J_{XY}(\nu_\lambda) < J_Y(\nu_\lambda)$ for all $0 < \lambda < 1$,

or

(ii) $\gamma^*(\mu_X, \mu_Y) = J_Y(\nu_\lambda) < J_{XY}(\nu_\lambda) < J_X(\nu_\lambda)$ for all $0 < \lambda < 1$.

Suppose that (i) holds. In particular, $J_X(\nu_\lambda)$ is finite and independent of $\lambda$ for $0 < \lambda < 1$. Therefore, by Lemma 5 (either part (a) or part (b)), there exists $\sigma \in M_1(\Sigma)$ with $J_X(\sigma) > J_X(\nu_\lambda) = \gamma^*(\mu_X, \mu_Y)$. Consider the probability measures $\bar{\nu}_\lambda = (1-\lambda)\nu_{1/2} + \lambda\sigma$. By the quasi-concavity of $J_X(\cdot)$ (see (16)), it follows that $J_X(\bar{\nu}_\lambda) > \gamma^*(\mu_X, \mu_Y)$ for all $0 < \lambda < 1$. Since $H(\sigma|\mu_X \times \mu_Y) < \infty$ for every $\sigma \in M_1(\Sigma)$ (by finiteness of $\Sigma$ and positivity of $\mu_X \times \mu_Y$), it is not hard to check the inequalities

$$\liminf_{\lambda \to 0} J_{XY}(\bar{\nu}_\lambda) \geq J_{XY}(\nu_{1/2}) > \gamma^*(\mu_X, \mu_Y)$$

and

$$\liminf_{\lambda \to 0} J_Y(\bar{\nu}_\lambda) \geq J_Y(\nu_{1/2}) > \gamma^*(\mu_X, \mu_Y).$$

Consequently, for $\lambda$ small enough, $J(\bar{\nu}_\lambda) > \gamma^*(\mu_X, \mu_Y)$ in contradiction with the definition of $\gamma^*(\mu_X, \mu_Y) = \sup_\nu J(\nu)$. The same argument applies to case (ii) above with the roles of $J_X(\cdot)$ and $J_Y(\cdot)$ interchanged. ∎

## 4. Sequences of different lengths.

We consider next the more general set-up with $n_X$ the length of the $\{X_i\}$ sequence and $n_Y$ the length of the $\{Y_i\}$ sequence, where $n_X$ and $n_Y$ can differ. Then, $M_n$ is replaced by

$$M_{n_X, n_Y} = \max_{\substack{0 \leq i \leq n_X - \Delta \\ 0 \leq j \leq n_Y - \Delta \\ \Delta \geq 0}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell})\},$$

and we define $\Delta^*_{n_X, n_Y}$, $\nu^*_{n_X, n_Y}$ and $M^U_{n_X, n_Y}$ corresponding to $\Delta^*_n, \nu^*_n$ and $M^U_n$, respectively. Suppose that $n_X \to \infty$ and $n_Y \to \infty$ in the manner that $\log n_X / \log(n_X n_Y) \to \lambda \in (0, 1)$. It is not hard to check that all the previous results obtained for $n_X = n_Y$ remain valid provided $\lambda = 1/2$, in particular accommodating the case of $n_X/n_Y$ bounded away from 0 and $\infty$.

The extension to $\lambda \neq 1/2$ is quite routine (cf. Arratia and Waterman (1985), Section 6). For completeness we state the counterparts of Theorems 1-5, omitting all proofs. The hypothesis (H) is unaltered. However, $H^*(\nu|\mu_X, \mu_Y)$ is now modified to the form

$$H^*(\nu|\mu_X, \mu_Y) = \frac{1}{2}\max\{H(\nu|\mu_X \times \mu_Y), H(\nu_X|\mu_X)/\lambda, H(\nu_Y|\mu_Y)/(1-\lambda)\}.$$

Letting $J(\nu) = E_\nu(F)/H^*(\nu)$, the definition of $\gamma^*(\mu_X, \mu_Y)$ is unchanged and $J_U$ remains as described in Theorem 3. Theorems 1,2 and 3 persist with $M_n/\log n$, $\Delta_n^*/\log n$ and $M_n^U/\log n$ replaced, respectively, by $M_{n_X,n_Y}/\log\sqrt{n_X n_Y}$, $\Delta^*_{n_X,n_Y}/\log\sqrt{n_X n_Y}$ and $M^U_{n_X,n_Y}/\log\sqrt{n_X n_Y}$. Both $\theta^*$ and $\alpha^*$ are unchanged and the necessary and sufficient condition for $\gamma^* = 2/\theta^*$ and $\mathcal{M} = \{\alpha^*\}$ is now

$$(\mathrm{E}_\lambda) \qquad H(\alpha^*|\mu_X \times \mu_Y) \geq \max\{H(\alpha^*_X|\mu_X)/\lambda, H(\alpha^*_Y|\mu_Y)/(1-\lambda)\}\ .$$

With regard to Theorem 4, even for the case of identical alphabets, symmetric scores ($F(y,x) = F(x,y)$) and $\mu_X = \mu_Y$, the condition $(E_\lambda)$ holds only for $\lambda \in [\lambda_{cr}, 1-\lambda_{cr}]$, with $\lambda_{cr} = 1/2$ when $F(x,y) = F(x) + F(y)$ and in general $\lambda_{cr} = H(\alpha^*_X|\mu_X)/\theta^* E_{\alpha^*}(F)$ satisfying $0 \leq \lambda_{cr} < 1/2$. The statement and the proof of Theorem 5 are unchanged.

**5. More than two sequences.**

Another extension concerns scores attendant to multiple, $r > 2$, independent sequences. Accordingly, consider $X_1^j, \cdots, X_{n_j}^j$ i.i.d. governed by the law $\mu_j \in M_1(\Sigma_j)$, respectively, and independent processes $\{X_i^j\}_{i=1}^\infty$. The scoring function $F$ now depends on a configuration of $r$ letters, one per sequence, i.e. $F : \Sigma = \prod_{j=1}^r \Sigma_j \to I\!R$, with

$$(\mathrm{H}) \qquad E_{\mu_1\times\mu_2\times\cdots\times\mu_r}(F) < 0, \quad (\mu_1 \times \mu_2 \times \cdots \times \mu_r)(F > 0) > 0.$$

The maximal segment score allowing shifts is now

$$M_{n_1,\cdots,n_r} = \max_{0\leq i_j \leq n_j-\Delta, j=1,\cdots,r} \{\sum_{\ell=1}^{\Delta} F(X^1_{i_1+\ell}, X^2_{i_2+\ell}, \cdots, X^r_{i_r+\ell})\}\ .$$

With $\log n_j/\log(n_1 n_2\cdots n_r) \to \lambda_j$, $\lambda_j \in (0,1)$, the relevant function $H^*(\nu|\mu_1,\ldots,\mu_r)$ becomes

$$H^*(\nu|\mu_1,\ldots,\mu_r) = \frac{1}{r}\max_{S\subset\{1,\cdots,r\}}\{H(\nu_S|\prod_{j\in S}\mu_j)/\sum_{j\in S}\lambda_j\}\ ,$$

where $\nu_S$ denotes the marginal of $\nu \in M_1(\Sigma)$ on the subspace $\prod_{j\in S}\Sigma_j$. With $J(\nu)$ and $\gamma^*$ defined following (1), Theorems 1, 2 and 3 carry over with the normalization factor being $\frac{1}{r}\log(n_1 n_2\cdots n_r)$.

The only proof whose extension may not be straightforward is Lemma 3. To ease the exposition we sketch the proof of Lemma 3 for the case $r = 3$. Fix $\nu$ and $k$ and let $N_j$, $j = 1,2,3$ be the number of segments of size $k$ with empirical measures $\nu_j$, $j = 1,2,3$ in the first, second and third sequence, respectively. Conditioned on $\{N_j\}_{j=1}^3$, let $B_{\ell_1\ell_2\ell_3}$ be the event that the $\ell_j$ segments among those with empirical measure $\nu_j$ in the $j$-th sequence, $j = 1,2,3$, are arranged such that their joint empirical measure is $\nu$. Now, pairs of these equally probable events are independent as long as they have at most one common index. Consider $W = \sum_{\ell_1=1}^{N_1}\sum_{\ell_2=1}^{N_2}\sum_{\ell_3=1}^{N_3} 1_{B_{\ell_1\ell_2\ell_3}}$ and designate $p = P(B_{\ell_1\ell_2\ell_3})$, $p_{12} = P(B_{112}|B_{111})$, $p_{13} = P(B_{121}|B_{111})$, and $p_{23} = P(B_{211}|B_{111})$. We have $E[W|\{N_j\}] = pN_1N_2N_3$ and

$$\mathrm{VAR}[W|\{N_j\}] \le pN_1N_2N_3\left[1 + (N_3-1)p_{12} + (N_2-1)p_{13} + (N_1-1)p_{23}\right].$$

By similar arguments as in the proof of Lemma 3 we deduce

$$P(W=0|\{N_j\}) \le \frac{1}{pN_1N_2N_3} + \frac{p_{12}}{pN_1N_2} + \frac{p_{13}}{pN_1N_3} + \frac{p_{23}}{pN_2N_3},$$

and

$$P(W=0) \le \frac{8k^3}{n^3pp_1p_2p_3} + \frac{4k^2p_{12}}{n^2pp_1p_2} + \frac{4k^2p_{13}}{n^2pp_1p_3} + \frac{4k^2p_{23}}{n^2pp_2p_3} + \frac{k}{np_1} + \frac{k}{np_2} + \frac{k}{np_3},$$

where $p_j = P(L_k^{\mathbf{X}^j} = \nu_j)$, $j = 1,2,3$. With the right hand side of (10) modified following the preceding inequality, the proof of the lemma is completed by observing that

$$\begin{aligned} pp_1p_2p_3 &= p_{12}p_3P(L_k^{(\mathbf{X}^1,\mathbf{X}^2)} = \nu_{\{1,2\}}) = p_{13}p_2P(L_k^{(\mathbf{X}^1,\mathbf{X}^3)} = \nu_{\{1,3\}}) = p_{23}p_1P(L_k^{(\mathbf{X}^2,\mathbf{X}^3)} = \nu_{\{2,3\}}) \\ &= P(L_k^{(\mathbf{X}^1,\mathbf{X}^2,\mathbf{X}^3)} = \nu). \end{aligned}$$

With $E_{\mu_1\times\cdots\times\mu_r}[e^{\theta^*F}] = 1$ and $\frac{d\alpha^*}{d(\mu_1\times\cdots\times\mu_r)} = e^{\theta^*F}$ the condition

$$(\mathrm{E}_\lambda)\qquad H(\alpha^*|\prod_{i=1}^r \mu_i) \ge \max_{S\subset\{1,\cdots,r\}} H(\alpha^*_S|\prod_{j\in S}\mu_j)/(\sum_{j\in S}\lambda_j),$$

is necessary and sufficient for $\gamma^* = r/\theta^*$, $\mathcal{M} = \{\alpha^*\}$.

In particular, $(E_\lambda)$ holds when $\lambda_j = 1/r$ for all $j$, all sequences have identical alphabets and identical laws, and the score $F$ is symmetric (i.e. invariant under permutations). To see this, observe that then $\alpha^*$ is invariant under permutations. Hence, $\alpha^*_S$ depends only on $|S| = k$ and condition $(E_\lambda)$ follows from the monotonicity in $k$ of $\tilde H(k) = \frac{1}{\binom{r}{k}}\sum_{|S|=k}\frac{1}{k}H(\alpha^*_S)$, that is $\tilde H(r) \le \tilde H(r-1) \le \cdots \le \tilde H(1)$, where $H(\alpha^*) = \sum_x \alpha^*(x)\log\frac{1}{\alpha^*(x)}$. For a proof of this well known property, see Han (1978) or Dembo et al. (1991).

**Remark:** In the case of $s$ identical alphabets ($\Sigma_j = \Sigma_1$, $j = 1,2,\ldots,s$), and $r$-letter scoring function $F : (\Sigma_1)^r \to I\!R$, many scoring systems focus on maximal intersequence segment score involving any subset of $r$ out of the $s$ given independent sequences, i.e.

$$M_{n_1,\cdots,n_s} = \max_{\substack{0\le i_j\le n_j-\Delta, j=1,\cdots,r \\ m_1\ne m_2\ne\cdots\ne m_r\in\{1,\ldots,s\}}} \{\sum_{\ell=1}^{\Delta} F(X^{m_1}_{i_1+\ell}, X^{m_2}_{i_2+\ell},\cdots,X^{m_r}_{i_r+\ell})\}.$$

Assuming that for each choice of distinct $m_j \in \{1, \dots, s\}$, $j = 1, \dots, r$, the measures $\mu_{m_1} \times \mu_{m_2} \times \cdots \times \mu_{m_r}$ satisfy hypothesis (H), the analysis carries over with $\gamma^*$ determined now as

$$\max_{m_1 \neq m_2 \neq \cdots \neq m_r \in \{1,\dots,s\}} \gamma^*(\mu_{m_1}, \mu_{m_2}, \dots, \mu_{m_r}) .$$

**6. Intrasequence maximal $r$-fold segment score.**

Suppose that $X_1, \cdots, X_n$ are i.i.d. $\mu_X$. With $\Sigma = \Sigma_X \times \Sigma_X$, a bounded score $F : \Sigma \to I\!R$ is considered, and the maximal (2-fold) segment score is of interest. Explicitly,

$$M_n = \max_{\substack{0 \leq i \neq j \leq n-\Delta \\ \Delta \geq 0}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, X_{j+\ell})\}. \tag{19}$$

We claim that Theorems 1–5 of Section 1 hold in this case, with $\mu_Y = \mu_X$. In particular, relevant to the strong laws, there is no difference in the intrasequence maximal segment score $M_n$ of (19) and the intersequence maximal segment score for i.i.d sequences. For the lower bound on $M_n$ (Lemmas 3 and 4) we simply divide the sequence $X_1, \cdots, X_n$ into two independent sequences of equal length $[n/2]$, assigning $Y_1 = X_{[n/2]+1}, \cdots, Y_{[n/2]} = X_n$. The maximal segment score allowing shifts between the latter two sequences, provides a lower bound on the intrasequence maximal score $M_n$ of (19). In fact, since $\log(n/2)/\log n \to 1$ as $n \to \infty$, this argument establishes (11) for the case at hand.

For the upper bound on $M_n$ (Lemmas 1 and 2) observe first that their validation in Section 2 relies only on the independence between the segments $(X_{i+1}, \cdots, X_{i+\Delta})$ and $(Y_{j+1}, \cdots, Y_{j+\Delta})$. Consequently, without any change they yield the correct upper bounds for the non-overlapping intrasequence score,

$$\widetilde{M}_n = \max_{\substack{0 \leq i,j \leq n-\Delta \\ |i-j| \geq \Delta \geq 0}} \{\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, X_{j+\ell})\}. \tag{20}$$

To treat the case of $1 \leq |i-j| < \Delta$, let us assume for definiteness that $i < j < i + \Delta$, and owing to the i.i.d. character of $\{X_i\}$ we may take $i = 0$, so we need to deal with $\sum_{\ell=1}^{\Delta} F(X_\ell, X_{j+\ell})$ for some $1 \leq j \leq (\Delta - 1)$.

For any $\Delta$ and any $j$, the following simple algorithm partitions the collection $\{(\ell, j+\ell)\}_{\ell=1}^{\Delta}$ into two disjoint subsets $S_1$ and $S_2$ differing in size by at most one, such that for each fixed $j+1 \leq a \leq \Delta$ the pairs $(a-j, a)$ and $(a, a+j)$ are not in the same subset. Thus, start with $(1, j+1) \in S_1$, then let $(j+1, 2j+1) \in S_2$, $(2j+1, 3j+1) \in S_1$, and continue in this alternating manner until reaching an index $kj + 1 > \Delta$. Now, if the most recent pair have been put in $S_1$, let $(2, j+2) \in S_2$ and vice versa. Continue in the manner described above until $kj + 2 > \Delta$, then move to $(3, j+3)$, and so on, until all pairs are exhausted.

Within the $\ell$ values for which $(\ell, j+\ell) \in S_1$ we may view $X_{j+\ell} = Y_\ell$ as samples from an independent second sequence of law $\mu_Y = \mu_X$. Similar treatment apply to the $\ell$ values for which $(\ell, j+\ell) \in S_2$.

With the union of the events

$$\{\sum_{\ell=1}^{\Delta} F(X_\ell, X_{j+\ell}) \geq 0\} \subset \{\sum_{(\ell,j+\ell)\in S_1} F(X_\ell, X_{j+\ell}) \geq 0\} \cup \{\sum_{(\ell,j+\ell)\in S_2} F(X_\ell, X_{j+\ell}) \geq 0\},$$

and with $|S_1|, |S_2| \geq \lfloor \Delta/2 \rfloor$, the proof of Lemma 1 follows increasing $c_0$ by a factor of 2.

Finally, in the proof of Lemma 2, the independence of the $\mathbf{X}$ and $\mathbf{Y}$ sequences is relevant only for $\nu \in \mathcal{L}_k$ such that

$$kH^*(\nu) = \frac{k}{2} H(\nu|\mu_X \times \mu_Y) \geq t \log n, \tag{21}$$

in which case (9) follows from the bound

$$P(A_{\nu,k}) \leq n^2 e^{-kH(\nu|\mu_X \times \mu_Y)} \leq n^{-2(t-1)} \leq n^{-(t-1)}.$$

This bound still holds for the non-overlapping choices of $i, j$, i.e. for

$$\bar{A}_{\nu,k} = \{\exists 0 \leq i, j \leq n-k, |i-j| > k, L_k^{(T^i X, T^j X)} = \nu\}.$$

There are at most $(2k-1)n$ choices of $i \neq j$ that cause an overlap between $(X_{i+1}, \cdots, X_{i+k})$ and $(X_{j+1}, \cdots, X_{j+k})$. For each of these choices we partition $\{(i+\ell, j+\ell)\}_{\ell=1}^k$ into the two subsets $S_1$ and $S_2$ as done above, and let $\nu_1, \nu_2$ be the candidate empirical measures of the pairs $(X_{i+\ell}, X_{j+\ell})$ with $(i+\ell, j+\ell) \in S_1$ and $(i+\ell, j+\ell) \in S_2$ respectively. Assuming with no real loss of generality that $k$ is even, $\nu = \frac{1}{2}(\nu_1 + \nu_2)$, and (21) implies that

$$\max\{\frac{k}{2} H(\nu_2|\mu_X \times \mu_X), \frac{k}{2} H(\nu_1|\mu_X \times \mu_X)\} \geq t \log n\ . \tag{22}$$

When considering the $k/2$ pairs within each $S_i$ by themselves, we may regard the relevant $X_\ell$ random variables as resulting from two independent sequences of length $k/2$ each. Hence, it follows that

$$P(A_{\nu,k} \setminus \bar{A}_{\nu,k}) \leq n(2k-1) \min\{P(L_{k/2}^{(\mathbf{X},\mathbf{Y})} = \nu_1), P(L_{k/2}^{(\mathbf{X},\mathbf{Y})} = \nu_2)\}.$$

Combining the basic bound (3) with (22) it thus follows that

$$P(A_{\nu,k}) \leq 2kn^{-(t-1)}.$$

With $k \leq c_0 \log n$, the above inequality suffices for the conclusion that for $J_U > 0$ and $t > 1$,

$$P(\bar{M}_n^U \geq t J_U \log n) \leq 2(c_0 \log n + 1)^{|\Sigma|+1} n^{-(t-1)}\ . \tag{23}$$

Coupled with Lemma 1 and Lemma 4 the bound (23) suffices for proving Theorem 3.

Similar arguments identify the results for maximal $r$-fold segment score

$$M_n^r = \max_{\substack{0\le i_1\neq i_2\neq\cdots\neq i_r\le n-\Delta \\ \Delta\ge 1}} \{\sum_{\ell=1}^{\Delta} F(X_{i_1+\ell}, X_{i_2+\ell}, \cdots, X_{i_r+\ell})\}, \quad r>2,$$

with the results of Section 5, where in this case $\Sigma_j = \Sigma_X, \mu_j = \mu_X$ and $\lambda_j = 1/r$ for $j = 1, \cdots, r$. In particular, when $F : (\Sigma_X)^r \to I\!R$ is invariant under permutations, $(E_\lambda)$ holds and thus,

$$M_n^r/\log n \to r/\theta^* \quad \text{and} \quad \nu_n^* \to \alpha^*$$

where $E_{\mu_X^r}[e^{\theta^* F}] = 1$ and $\frac{d\alpha^*}{d\mu_X^r} = e^{\theta^* F}$.

**7. Asymptotics of the longest quality $q$ segment.**

As in Section 1, let $X_1, \cdots, X_n$ be i.i.d. $\mu_X$ and $Y_1, \cdots, Y_n$ be i.i.d. $\mu_Y$. We are now interested in characterizing all sets of pairs $(x, y) \in S \subset \Sigma$, and segments yielding a large relative occupation time of $S$. Formally, a threshold level $q > (\mu_X \times \mu_Y)(S) > 0$ is specified and a segment $\{(X_{i+\ell}, Y_{j+\ell})\}_{\ell=1}^{\Delta}$ is said to be of quality $q$ if

$$\sum_{\ell=1}^{\Delta} 1_S(X_{i+\ell}, Y_{j+\ell}) \ge q\Delta.$$

For fixed $S$, let $Q_n$ denote the maximal length among quality $q$ segments, i.e.,

$$Q_n = \max\{\Delta : \exists 0 \le i, j \le n-\Delta \text{ such that } \sum_{\ell=1}^{\Delta} 1_S(X_{i+\ell}, Y_{j+\ell}) \ge q\Delta\}.$$

This problem has been considered by Arratia and Waterman (1989) and Arratia et al. (1990) for $\Sigma_X = \Sigma_Y$ and $S = \{(x, x) : x \in \Sigma_X\}$. Without shifts the maximal length of quality $q$ segments, defined as

$$\widetilde{R}_n = \max\{\Delta : \exists 0 \le i \le n-\Delta \text{ such that } \sum_{\ell=1}^{\Delta} 1_S(X_{i+\ell}, Y_{i+\ell}) \ge q\Delta\},$$

is known to behave as $H(q|p)\log n + o(\log n)$, where $p = (\mu_X \times \mu_Y)(S)$ and $H(q|p) = q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}$ is the binary relative entropy, while the empirical measure $\nu_n^*$ within the longest quality $q$ segment converges to the measure $\alpha^*$, defined via

$$\frac{d\alpha^*}{d(\mu_X \times \mu_Y)} = \begin{cases} \frac{q}{p} & \text{on } S \\ \frac{1-q}{1-p} & \text{on } S^c \end{cases}$$

These are easy applications of the Erdös–Rényi strong law for coin tossing (cf. Erdös–Rényi (1970)). More finer asymptotic analysis of $\widetilde{R}_n$ is carried out in Arratia et al. (1990), using the Chen-Stein method for Poisson approximation.

To put the study of the asymptotics of $Q_n$ within the framework developed earlier, we define

$$J(\nu) = \begin{cases} 1/H^*(\nu) & \nu(S) \geq q \\ 0 & \text{otherwise,} \end{cases}$$

and let $\gamma^*(\mu_X, \mu_Y) = \max_\nu J(\nu)$ and $\mathcal{M} = \{\nu : J(\nu) = \gamma^*\}$, where $H^*(\nu) = H^*(\nu|\mu_X, \mu_Y)$ is as defined in Section 1. We then obtain the following analogs of Theorems 1 and 2.

**Theorem 1′.** *$\gamma^*(\mu_X, \mu_Y)$ is finite and positive and $Q_n/\log n \to \gamma^*(\mu_X, \mu_Y)$ a.s.*

**Theorem 2′.** *All limit points of $\nu_n^*$ belong to the set $\mathcal{M}$, a.s.*

In duality with hypothesis (H), our assumption that $q > (\mu_X \times \mu_Y)(S) > 0$ implies that $\gamma^*$ is finite and positive, the sets $\{\nu : J(\nu) \geq \beta\}$ are compact for all $\beta > 0$, and so is the non-empty set $\mathcal{M}$. Therefore, the preceding theorems are direct consequences of the following analog of Theorem 3.

**Theorem 3′.** *Let $M_F(\Sigma) = \{\nu : \nu(S) \geq q\}$. For any $U \subset M_F(\Sigma)$ let $J_U = \sup_{\nu \in U} J(\nu)$ and*

$$Q_n^U = \max\{\Delta : \exists 0 \leq i,j \leq n-\Delta, \text{ such that } \sum_{\ell=1}^{\Delta} 1_S(X_{i+\ell}, Y_{j+\ell}) \geq q\Delta\,,\ L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})} \in U\}$$

*where $L_\Delta^{(T^i\mathbf{X}, T^j\mathbf{Y})}$ is as defined in Theorem 3. Then, a.s.,*

$$J_{U^o} \leq \liminf_{n\to\infty} Q_n^U / \log n \leq \limsup_{n\to\infty} Q_n^U / \log n \leq J_U,$$

*where $U^o$ denotes the relative interior of $U$ in $M_F(\Sigma)$.*

The proof of Theorem 3′ is quite similar to the proof of Theorem 3 as presented in Section 2. We omit the details.

With $\alpha^*$, $\gamma^*$ and $\mathcal{M}$ as defined earlier, Theorems 4 and 5 apply to the present context if we define $F(x,y) = 1_S(x,y) - q$ (i.e. $\nu(S) \geq q$ iff $E_\nu(F) \geq 0$), and set $\theta^* = 1/H(q|p)$ (this definition of $\theta^*$ differs from the definition used in Sections 1-6). The proofs parallel those presented in Section 3, thus the details are omitted.

Specializing Theorems 1 and 4 to $\Sigma_X = \Sigma_Y$, $\mu_X = \mu_Y$ and $S = \{(x,x) : x \in \Sigma_X\}$, we recover Theorem 1 of Arratia and Waterman (1989).

In the framework of bounded vector-valued scores $F : \Sigma \to I\!R^d$, longest quality segments with shifts are defined by

$$Q_n^A = \max\{\Delta : \exists 0 \leq i,j \leq n-\Delta \text{ such that } \frac{1}{\Delta}\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell}) \in A\},$$

and without shifts by

$$\widetilde{R}_n^A = \max\{\Delta : \exists 0 \leq i \leq n-\Delta \text{ such that } \frac{1}{\Delta}\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{i+\ell}) \in A\},$$

where $A \subset \mathbb{R}^d$ is a closed subset of the convex hull of the support of $F(X_1, Y_1)$, such that $E_{\mu_X \times \mu_Y}[F(X,Y)] \notin A$, and $A = \overline{A^o}$. The preceding results correspond to $F(x,y) = 1_S(x,y) - q$ and $A = [0, 1-q]$. The asymptotics of $\widetilde{R}_n^A$ are treated in Dembo and Zeitouni (1993), Section 3.2. The formula determining $\gamma^*(\mu_X, \mu_Y)$ and $\nu^*$ for $Q_n^A$, is based on

$$J(\nu) = \begin{cases} 1/H^*(\nu) & E_\nu(F) \in A, \\ 0 & \text{otherwise,} \end{cases}$$

and for the analysis of Theorem 4 to apply, we should take

$$\theta^* = 1/ \inf_{\{\nu : E_\nu(F) \in A\}} H(\nu|\mu_X \times \mu_Y) ,$$

and set $\alpha^*$ to be the measure for which the infimum is achieved (this measure is unique for convex $A$).

Similar extension of the maximal segment score without shifts $R_n$ (of Section 1) to vector-valued scoring system is considered in Dembo and Zeitouni (1993), Section 5.5. Analogously, the maximal segment score allowing shifts will be defined as

$$M_n^A = \max_{\substack{0 \le i,j \le n-\Delta \\ \Delta \ge 0}} \{ y : \frac{1}{y} \sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell}) \in A \} ,$$

where $A = \overline{A^o} \subset \mathbb{R}^d$ is such that $tE_{\mu_X \times \mu_Y}[F(X,Y)] \notin A$, for all $t \ge 0$. Restricting attention to bounded scores and convex $A$ which excludes a cone around the ray $\{tE_{\mu_X \times \mu_Y}[F(X,Y)]\}_{t \ge 0}$, similar results hold for $M_n^A$ with the formula determining $\gamma^*(\mu_X, \mu_Y)$ and $\nu^*$ based on

$$J(\nu) = 1/ \inf_{\{t : t > 0, tE_\nu(F) \in A\}} tH^*(\nu) .$$

**8. Extension to Polish alphabets.**

Suppose now that $X_1, \cdots, X_n$ are i.i.d. $\mu_X$ and $Y_1, \cdots, Y_n$ are i.i.d. $\mu_Y$, where $\mu_X$ and $\mu_Y$ are Borel measures on Polish (complete, separable, metric) spaces $(\Sigma_X, d_X)$ and $(\Sigma_Y, d_Y)$ respectively. One motivating example is $\Sigma_X = \mathbb{R}^d$ and $\Sigma_Y$ a singleton, corresponding to the strong laws of Dembo and Karlin (1991a). Throughout, let $\Sigma = \Sigma_X \times \Sigma_Y$ denote the product Polish space, $M_1(\Sigma_X), M_1(\Sigma_Y), M_1(\Sigma)$ denote the sets of Borel probability measures on $\Sigma_X$, $\Sigma_Y$ and $\Sigma$ respectively, and let the score $F : \Sigma \to \mathbb{R}$ be any bounded, Borel measurable function on $\Sigma$. Henceforth, the Borel $\sigma$-fields on $\Sigma_X, \Sigma_Y$ and $\Sigma$ are taken as the completed $\sigma$-fields with respect to $\mu_X, \mu_Y$ and $\mu_X \times \mu_Y$, respectively, and $P$ denotes the product measure on $\Sigma^{\mathbb{N}}$ generated by $\mu_X \times \mu_Y$.

The random variables $M_n, \Delta_n^*$ and $\nu_n^*$ are now defined paralleling the finite alphabet case. The relative entropy is given by the formula

$$H(\nu|\mu) = \begin{cases} \int_\Sigma f \log f d\mu & \frac{d\nu}{d\mu} = f, \text{ exists} \\ \infty & \text{otherwise} \end{cases}$$

and (H) and $J(\nu)$ are as previously defined in Section 1. Since the relative entropy is a lower semi-continuous function in the $\tau$-topology on $M_1(\Sigma)$ having compact level sets ($\{\nu : H(\nu|\mu) \leq \alpha\}, \alpha < \infty$, see Deuschel and Stroock (1989), Section 3.2, or Dembo and Zeitouni (1993), Section 6.2), it follows by the continuity of the map $\nu \mapsto E_\nu(F)$, that the function $J(\cdot)$ is upper semicontinuous. Moreover, the upper boundedness of $F$ implies that the sets $\{\nu : J(\nu) \geq \beta\}$ are compact for all $\beta > 0$. Hence, $\gamma^*(\mu_X, \mu_Y)$ determined by (1) is finite and positive, and $\mathcal{M} = \{\nu \in M_1(\Sigma) : \gamma^*(\mu_X, \mu_Y) = J(\nu)\}$ is a non-empty compact set. We claim that both Theorems 1 and 2 hold in the present context, where the convergence of the empirical measures $\nu_n^*$ of Theorem 2 is now to be interpreted in the sense of weak convergence in $M_1(\Sigma)$. The proof of these results is based on a reduction to the finite alphabet case via the following approximation lemma.

*Lemma 6*. There exist $\{F_m\}_{m=1}^\infty$ such that:

(a) $\lim_{m\to\infty} \int_\Sigma |F - F_m| d\mu_X \times \mu_Y = 0$.

(b) $F_m(x,y) = \sum_{i=0}^{N_X^m} \sum_{j=0}^{N_Y^m} \alpha_{ij}^m 1_{A_i^m}(x) 1_{B_j^m}(y)$, where $\{A_i^m\}_{i=0}^{N_X^m}$ is a finite partition of $\Sigma_X$ to disjoint measurable sets, $\{B_j^m\}_{j=0}^{N_Y^m}$ is a finite partition of $\Sigma_Y$ to disjoint measurable sets, which may be ordered by refinements, and $|\alpha_{ij}^m| \leq \|F\|_\infty$.

Proof: Since $C_b(\Sigma)$, the collection of bounded, continuous, real-valued functions on $\Sigma$ is dense in $L_1(\Sigma, \mu_X \times \mu_Y)$, it suffices to consider $F \in C_b(\Sigma)$. With $\mu_X$ tight in $\Sigma_X$, and $\mu_Y$ tight in $\Sigma_Y$ there exist compacts $K_X^m \subset \Sigma_X$ and $K_Y^m \subset \Sigma_Y$ such that $\mu_X(K_X^m) \geq 1 - 1/m$ and $\mu_Y(K_Y^m) \geq 1 - 1/m$.

Since $F$ is uniformly continuous on the compact set $K_X^m \times K_Y^m$, there exist $\delta_m > 0$ small enough such that, for all $(x,y) \in K_X^m \times K_Y^m$, $(x',y') \in K_X^m \times K_Y^m$,

$$d_X(x,x') \leq \delta_m, d_Y(y,y') \leq \delta_m \Rightarrow |F(x,y) - F(x',y')| \leq 1/m.$$

Moreover, there exist finite partitions $\{A_i^m\}_{i=1}^{N_X^m}$ of $K_X^m$ and $\{B_j^m\}_{j=1}^{N_Y^m}$ of $K_Y^m$ such that

$$\operatorname{diam}(A_i^m) \leq \delta_m, \quad \text{and} \quad \operatorname{diam}(B_j^m) \leq \delta_m,$$

where $\operatorname{diam}(A) = \sup_{x,x' \in A} d_X(x,x')$. Taking $A_0^m = (K_X^m)^c$, $B_0^m = (K_Y^m)^c$ and $\alpha_{ij}^m = F(x_i, y_j)$ for an arbitrary choice $x_i \in A_i^m, y_j \in B_j^m$, it follows that

$$\begin{aligned}\int_\Sigma |F - F_m| d\mu_X \times \mu_Y &\leq \sum_{i=0}^{N_X^m} \sum_{j=0}^{N_Y^m} \int_{A_i^m \times B_j^m} |F - F(x_i,y_j)| d\mu_X \times \mu_Y \\ &\leq 2\|F\|_\infty [\mu_X(A_0^m) + \mu_Y(B_0^m)] + 1/m \leq (4\|F\|_\infty + 1)/m.\end{aligned}$$

Consequently, both (a) and (b) hold for this sequence of functions. ∎

To facilitate the use of Lemma 6, we make the following observation.

Lemma 7. If $\int_\Sigma |F - F_m| d\mu_X \times \mu_Y \to 0$ and $\|F_m\|_\infty \le \|F\|_\infty$ then:

(a) For every $\delta > 0$, there exist $\eta_m(\delta) \to \infty$, as $m \to \infty$ such that for all $k \ge 1$,

$$P(\sum_{\ell=1}^{k} |F(X_\ell, Y_\ell) - F_m(X_\ell, Y_\ell)| \ge k\delta) \le e^{-k\eta_m(\delta)}$$

(b) For every $\alpha < \infty$, $\int_\Sigma |F - F_m| d\nu \to 0$ as $m \to \infty$ uniformly in $\{\nu : H(\nu|\mu_X \times \mu_Y) \le \alpha\}$.

Proof: (a) By Chebychev's bound it follows that for every $\delta > 0$ and integer $k, m$

$$\frac{1}{k} \log P(\sum_{\ell=1}^{k} |F(X_\ell, Y_\ell) - F_m(X_\ell, Y_\ell)| \ge k\delta) \le -\Lambda_m^*(\delta) \ ,$$

where $\Lambda_m^*(\cdot)$ is the Fenchel-Legendre transform of $\Lambda_m(\lambda) = \log E_{\mu_X \times \mu_Y}[e^{\lambda|F-F_m|}]$. With $|F - F_m| \le 2\|F\|_\infty$ and $E_{\mu_X \times \mu_Y}[|F - F_m|] \to 0$ as $m \to \infty$, it is easy to check that for every fixed $\delta > 0$ $\Lambda_m^*(\delta) \to \infty$ as $m \to \infty$.
(b) Fix $\alpha < \infty$ and $\nu$ such that $H(\nu|\mu_X \times \mu_Y) \le \alpha$. Then, with $f = \frac{d\nu}{d(\mu_X \times \mu_Y)}$,

$$\int_\Sigma |F - F_m| d\nu \le M \int_\Sigma |F - F_m| d\mu_X \times \mu_Y + 2\|F\|_\infty \int_\Sigma 1_{\{f \ge M\}} \, d\nu \ .$$

Recall that

$$\begin{aligned} \log M \int_\Sigma 1_{\{f \ge M\}} d\nu - 1/e &\le \int_\Sigma (\log f \vee 0) d\nu + \inf_{x \ge 0} [x \log x] \int_\Sigma d\mu_X \times \mu_Y \\ &\le \int_\Sigma f \log f \, d\mu_X \times \mu_Y = H(\nu|\mu_X \times \mu_Y) \le \alpha, \end{aligned}$$

implying that

$$\int_\Sigma 1_{\{f \ge M\}} d\nu \le (\alpha + 1/e)/\log M$$

and thus completing the proof. ∎

With the assistance of Lemmas 6 and 7, we turn to prove Theorem 1 in the present context. First observe that Lemma 1 applies even though the alphabet is no longer finite. Consequently, it suffices to consider $\bar{M}_n$, the maximal segment score, restricting to $\Delta \le c_0 \log n$ with $c_0$ as determined by Lemma 1. With $\bar{M}_n^m$ denoting the analog maximal segment score corresponding to the approximations $\{F_m\}$ of Lemma 6, we have that the event $\{|\bar{M}_n - \bar{M}_n^m| \ge \delta \log n\}$ is contained in the event

$$\{\exists i, j \le n - \Delta, \Delta \le c_0 \log n, \ |\sum_{\ell=1}^{\Delta} F(X_{i+\ell}, Y_{j+\ell}) - \sum_{\ell=1}^{\Delta} F_m(X_{i+\ell}, Y_{j+\ell})| \ge \delta \log n\}.$$

Hence, by the union of events bound

$$P(|\bar{M}_n - \bar{M}_n^m| \geq \delta \log n) \leq n^3 \sup_{1 \leq \Delta \leq c_0 \log n} P(\sum_{\ell=1}^{\Delta} |F(X_\ell, Y_\ell) - F_m(X_\ell, Y_\ell)| \geq \delta \log n)$$
$$\leq n^3 e^{-(c_0 \log n)\eta_m(\delta/c_0)},$$

where the second inequality follows by part (a) of Lemma 7. With $\eta_m(\delta/c_0) \to \infty$ as $m \to \infty$ it thus follows that almost surely, for all $m \geq m_0(\delta)$,

$$\lim_{n\to\infty} |\frac{\bar{M}_n}{\log n} - \frac{\bar{M}_n^m}{\log n}| \leq \delta. \tag{24}$$

Assuming that (H) holds, it easily follows that for all $m$ large enough

$$E_{\mu_X \times \mu_Y}[F_m] < 0, \quad \mu_X \times \mu_Y(F_m > 0) > 0\ . \tag{H$_\mathrm{m}$}$$

Consequently, Theorem 1 applies to the asymptotics of $\bar{M}_n^m$ for all $m$ large enough (since the partitions $\{A^m\}$ and $\{B^m\}$ map $\Sigma_X$ and $\Sigma_Y$ respectively into finite alphabets that suffice for representing the joint law of $\{F_m(X_i, Y_j)\}$).

Letting $\gamma_m^*(\mu_X, \mu_Y)$ denote the corresponding limits of $\bar{M}_n^m/\log n$, the proof of Theorem 1 in the current set-up is completed by showing that

$$\lim_{m\to\infty} \gamma_m^*(\mu_X, \mu_Y) = \gamma^*(\mu_X, \mu_Y). \tag{25}$$

To this end observe first that

$$\gamma_m^*(\mu_X, \mu_Y) = \sup_{\nu \in M_{F_m}(\Sigma)} J_m(\nu)$$

where

$$J_m(\nu) = E_\nu(F_m)/H^*(\nu|\mu_X \times \mu_Y)\ .$$

Let now $\nu_m$ be such that $J_m(\nu_m) = \gamma_m^*(\mu_X, \mu_Y)$ and $\nu^*$ be such that $J(\nu^*) = \gamma^*(\mu_X, \mu_Y)$ (such measures exist as soon as (H) and (H$_\mathrm{m}$) hold, paraphrasing the discussion below (1)).

Since $J_m(\nu_m) \geq J_m(\nu^*)$ it follows that

$$\liminf_{m\to\infty} \gamma_m^*(\mu_X, \mu_Y) \geq \liminf_{m\to\infty} E_{\nu^*}(F_m)/H^*(\nu^*),$$

Obviously, since $\gamma^*(\mu_X, \mu_Y) = E_{\nu^*}(F)/H^*(\nu^*) > 0$ it follows that $\infty > H(\nu^*|\mu_X \times \mu_Y) > 0$, and hence by part (b) of Lemma 7, $E_{\nu^*}(F_m) \to E_{\nu^*}(F)$, yielding the inequality

$$\liminf_{m\to\infty} \gamma_m^*(\mu_X, \mu_Y) \geq \gamma^*(\mu_X, \mu_Y) \tag{26}$$

Note that with $E_{\nu_m}(F_m) \leq \|F\|_\infty$, (26) implies that $H^*(\nu_m)$ is bounded uniformly in $m$ and hence so is $H(\nu_m|\mu_X \times \mu_Y)$. Consequently, $|E_{\nu_m}(F_m) - E_{\nu_m}(F)| \to 0$ by part (b) of Lemma 7, and as soon as we show that $H^*(\nu_m)$ are bounded away from zero we will have

$$|\gamma_m^*(\mu_X, \mu_Y) - J(\nu_m)| \to 0,$$

implying that (25) holds (as $\gamma^*(\mu_X, \mu_Y) \geq J(\nu_m)$).

Suppose now that $H^*(\nu_m)$ are not bounded away from zero. Then, by the compactness of level sets of $H(\cdot|\mu_X \times \mu_Y)$ in the $\tau$-topology of $M_1(\Sigma)$, passing to a subsequence, $H^*(\nu_m) \to 0$ yields $H(\nu_m|\mu_X \times \mu_Y) \to 0$ resulting with $\nu_m \to \mu_X \times \mu_Y$ in $\tau$-topology of $M_1(\Sigma)$. In particular $E_{\nu_m}(F) \to E_{\mu_X \times \mu_Y}(F)$, implying by (H) that for some $\delta > 0$ and all $m$ large enough $E_{\nu_m}(F) \leq -\delta$. With $|E_{\nu_m}(F_m) - E_{\nu_m}(F)| \to 0$, (again by part (b) of Lemma 7), it follows that $E_{\nu_m}(F_m) < 0$ for all $m$ large enough contradicting the underlying assumption that $J_m(\nu_m) = \gamma_m^*(\mu_X, \mu_Y) > 0$. In conclusion, $H^*(\nu_m)$ are bounded away from zero, hence (25) holds, and the proof of Theorem 1 is completed.

Utilizing the same reduction to finite alphabets, and applying there Theorem 3, Theorem 2 holds in the current set-up provided that for all $\delta > 0$

$$\gamma^*(\mu_X, \mu_Y) > \liminf_{m\to\infty} \sup_{\left\{\substack{\nu(A_i^m \times B_j^m) = \nu'(A_i^m \times B_j^m) \\ d(\nu', \mathcal{M}) \geq \delta}\right\}} J_m(\nu),$$

where here $d(\cdot, \mathcal{M})$ denotes the Lévy metric compatible with weak convergence, and $\{A_i^m\}, \{B_j^m\}$ are the partitions specified in Lemma 6.

To this end, assume that

$$\gamma^*(\mu_X, \mu_Y) \leq \liminf_{m\to\infty} \sup_{\left\{\substack{\nu(A_i^m \times B_j^m) = \nu'(A_i^m \times B_j^m) \\ d(\nu', \mathcal{M}) \geq \delta}\right\}} J_m(\nu),$$

i.e., there exist $\nu_m, \nu'_m$ such that

$$\liminf_{m\to\infty} J_m(\nu_m) \geq \gamma^*$$

and

$$\nu_m(A_i^m \times B_j^m) = \nu'_m(A_i^m \times B_j^m),\ d(\nu'_m, \mathcal{M}) \geq \delta\ . \tag{27}$$

Note that for large enough $m$, $H(\nu_m|\mu_X \times \mu_Y) \leq 3||F||_\infty/\gamma^*$, hence passing to a suitable subsequence, $\nu_m$ converges in the $\tau$ topology to some $\nu^*$. By part (b) of Lemma 7 and the upper semicontinuity of $J(\cdot)$, it follows that $J(\nu^*) \geq \gamma^*$, and hence $\nu^* \in \mathcal{M}$.

We now claim that necessarily, $\nu'_m$ converges weakly to $\nu^*$, contradicting (27). Indeed, note that $\{A_i^m\}$ and $\{B_j^m\}$ could be taken as forming a refinement (in $m$) of partitions. Then, for each $i, j, m_0$,

$$\nu'_m(A_i^{m_0} \times B_j^{m_0}) \to \nu^*(A_i^{m_0} \times B_j^{m_0})\ .$$

Let now $G$ be an arbitrary open subset of $\Sigma$. Then, up to the $\mu_X \times \mu_Y$ null set $(\cup_m K_X^m \times K_Y^m)^c$, the set $G$ may be written as a countable union of terms of the form $A_{i_k}^{m_k} \times B_{j_k}^{m_k}$. Because $A_i^m \times B_j^m$ are disjoint for every fixed $m$, and form a refinement with respect to $m$, every finite union of such terms can be made into a finite union of disjoint sets of this form. Thus, using Fatou's lemma, and the absolute continuity of $\nu^*$ with respect to $\mu_X \times \mu_Y$,

$$\liminf_{m\to\infty} \nu'_m(G) \geq \nu^*(G) \ ,$$

which yields the required weak convergence and hence the contradiction.

**Remarks:** (a) For $\Sigma_X = \Sigma_Y$ and $\mu_X = \mu_Y$, the construction of Lemma 6 allows for $\{A^m\} = \{B^m\}$ for all $m$. Thus, the preceding arguments apply for the non-overlapping intrasequence score $\widetilde{M}_n$ of (20), for which Theorems 1 and 2 persist in the present context of $X_i$ i.i.d. $\mu_X$, taking values in the Polish space $(\Sigma_X, d_X)$. By splitting the overlapping pairs into index sets $S_1$, $S_2$ of almost equal size before applying part (a) of Lemma 7, we conclude that (24) applies also for the intrasequence score $M_n$ of (19), and consequently, Theorems 1 and 2 persist in the context of $M_n$ of (19) and of $X_i$ taking values in a Polish space.

(b) For $S \subset \Sigma$ Borel measurable, the approach taken earlier can be modified as to allow for the extension of Theorems $1'$ and $2'$ of Section 7 to the present context of general Polish alphabets. To this end, Lemmas 6 and 7 are replaced by the following standard approximation lemmas whose proof is omitted.

_Lemma 6′._ There exist $S_m^- \subset S \subset S_m^+$ measurable sets of $\Sigma$, with $S_m^+$ monotone decreasing in $m$, $S_m^-$ monotone increasing in $m$, such that

(a) $\lim_{m\to\infty} \mu_X \times \mu_Y(S_m^+ \setminus S) = \lim_{m\to\infty} \mu_X \times \mu_Y(S \setminus S_m^-) = 0$.

(b) For each $m$, both $S_m^+$ and $S_m^-$ are finite unions of disjoint product sets of the form $A_i^m \times B_j^m$ with measurable $A_i^m \subset \Sigma_X$ and $B_j^m \subset \Sigma_Y$, and the partitions induced on $\Sigma_X$ and $\Sigma_Y$ by $\{A_i^m\}_i$ and $\{B_j^m\}_j$ may be ordered by refinements.

_Lemma 7′._ For every $\beta > 0$, both $\nu(S_m^+) \searrow \nu(S)$ and $\nu(S_m^-) \nearrow \nu(S)$ as $m \to \infty$, uniformly in $\{\nu : J(\nu) \geq \beta\}$.

To the sets $S_m^-$ $(S_m^+)$ correspond longest quality $q$ segments in finite alphabets, induced by the relevant partitions, say $Q_n^{m^-}$ ($Q_n^{m^+}$ respectively). The set inclusions in Lemma $6'$ imply that $Q_n^{m^-} \leq Q_n \leq Q_n^{m^+}$. Hence, with Theorems $1'$, $3'$ already holding for finite alphabets, the general case follows provided that

$$\lim_{m\to\infty} \inf_{\{\nu:\nu(S_m^-)\geq q\}} H^*(\nu) \leq \inf_{\{\nu:\nu(S)\geq q\}} H^*(\nu) \leq \lim_{m\to\infty} \inf_{\{\nu:\nu(S_m^+)\geq q\}} H^*(\nu) \tag{28}$$

and for all $\delta > 0$

$$\limsup_{m\to\infty} \inf_{\left\{ \substack{\nu(A_i^m \times B_j^m) = \nu'(A_i^m \times B_j^m) \\ d(\nu',\mathcal{M}) \geq \delta, \nu(S_m^+) \geq q} \right\}} H^*(\nu) > \inf_{\{\nu:\nu(S)\geq q\}} H^*(\nu) \tag{29}$$

Applying Lemma 7′ and paraphrasing the arguments presented earlier, it is easy to check that (29) holds, while (28) holds provided that $f(a) = \inf_{\{\nu:\nu(S)\geq a\}} H^*(\nu)$ is continuous for all $a > p$. The convexity of $H^*(\nu)$ and the compactness of its level sets imply that $f(a)$ is convex, and being finite on $a \in [p, 1]$, it is also continuous on $(p, 1)$. The continuity of $f(\cdot)$ at $a = 1$ can be verified directly.

The definition of $\theta^*$ and $\alpha^*$ is exactly as in the finite alphabet case and then the analogs of Theorems 4 and 5 are:

**Theorem 4′.** *$\gamma^*(\mu_X, \mu_Y) \leq 2/\theta^*$ and $\gamma^*(\mu_X, \mu_Y) = 2/\theta^*$ iff (E) holds in which case $\mathcal{M} = \{\alpha^*\}$. In particular (E) holds whenever $\Sigma_X = \Sigma_Y$, $F(x, y) = F(y, x)$ and $\mu_X = \mu_Y$.*

Proof. Apart from (13) the proof of these statements in Section 3 did not rely on finiteness of $\Sigma_X$ or $\Sigma_Y$. For completeness, we outline here a proof of (13) in the current set-up. First, since the relative entropy function is non-negative and monotone with respect to partitions, the left hand side of (13) is infinite as soon as the right hand side is infinite. Hence, we may assume that $f = \frac{d\nu}{d\mu_X \times \mu_Y}$ exists and let $f_X = \int_{\Sigma_Y} f d\mu_Y$ and $f_Y = \int_{\Sigma_X} f d\mu_X$ denote the marginals of $\frac{d\nu}{d\mu_X \times \mu_Y}$, in which case

$$H(\nu_X|\mu_X) + H(\nu_Y|\mu_Y) - H(\nu|\mu_X \times \mu_Y) = \int_{\{f>0\}} \log(\frac{f_X f_Y}{f}) d\nu \ .$$

The proof is complete by applying Jensen's inequality and checking that $\int_{\{f>0\}} \frac{f_X f_Y}{f} d\nu \leq 1$. ∎

**Theorem 5′.** *Let*

$$G_X(x) = \operatorname{ess}\sup_{y\in\Sigma_Y} F(x, y)$$

*with the essential supremum being with respect to $\mu_Y$ and defined up to $\mu_X$ null sets of values of $x$. Similarly, let*

$$G_Y(y) = \operatorname{ess}\sup_{x\in\Sigma_X} F(x, y).$$

*If $E_{\mu_X}(G_X) \geq 0$ and $E_{\mu_Y}(G_Y) \geq 0$ then $\mathcal{M} = \{\nu^*\}$.*

**Remark.** Note that by an application of Lusin's theorem both $G_X$ and $G_Y$ are measurable with respect to the completed Borel $\sigma$-fields of $\Sigma_X$ and $\Sigma_Y$, respectively.

Proof. The proof of uniqueness of $\nu^*$ as presented in Section 3 carries over as soon as we show that there exist $\sigma \in M_1(\Sigma)$ with $H(\sigma|\mu_X \times \mu_Y) < \infty$ and with $J_X(\sigma)$ (or with $J_Y(\sigma)$) arbitrarily large. Thus, all that is needed here is to adapt the argument of part (a) of Lemma 5, using the following well known result:

$$E_{\mu_X}[G_X] = \lim_{\lambda\to\infty} E_{\mu_X}[G_X^\lambda] \tag{30}$$

with

$$G_X^\lambda(x) = \frac{\int_{\Sigma_Y} F(x, y)e^{\lambda F(x,y)} d\mu_Y}{\int_{\Sigma_Y} e^{\lambda F(x,y)} d\mu_Y} \ .$$

Let $\sigma_\lambda$ be defined by

$$\frac{d\sigma_\lambda}{d(\mu_X \times \mu_Y)} = \frac{e^{\lambda F(x,y)}}{\int_{\Sigma_Y} e^{\lambda F(x,y)} d\mu_Y} \, .$$

Note that $H(\sigma_\lambda|\mu_X \times \mu_Y) \le 2\lambda\|F\|_\infty < \infty$. Moreover, for all $\lambda$, $(\sigma_\lambda)_X = \mu_X$ and hence $H((\sigma_\lambda)_X|\mu_X) = 0$. Consequently, if $E_{\mu_X}[G_X] > 0$ then by (30), $E_{\sigma_\lambda}(F) = E_{\mu_X}[G_X^\lambda] > 0$ for $\lambda$ large enough implying that $J_X(\sigma_\lambda) = \infty$ while $H(\sigma_\lambda|\mu_X \times \mu_Y) < \infty$ as required.

A slightly more involved construction is needed when $E_{\mu_X}[G_X] = 0$. In this case let $a = \mu_X \times \mu_Y(F > 0) > 0$ and define $\phi \in M_1(\Sigma)$ via

$$\frac{d\phi}{d(\mu_X \times \mu_Y)} = \frac{1_{\{F>0\}}}{a} = f \, .$$

Since $f$ is bounded, $H(\phi|\mu_X \times \mu_Y) < \infty$. Moreover,

$$E_\phi(F) = E_{\mu_X \times \mu_Y}[F 1_{\{F>0\}}]/a = 2\delta > 0 \, .$$

By (30), for every $\epsilon > 0$ there exists $\lambda_\epsilon$ large enough for $E_{\mu_X}[G_X^{\lambda_\epsilon}] \ge -\epsilon\delta$. Consider now the probability measures $\phi_\epsilon = (1-\epsilon)\sigma_{\lambda_\epsilon} + \epsilon\phi$. Note that $H(\phi_\epsilon|\mu_X \times \mu_Y) < \infty$, and by the above choice of $\lambda_\epsilon$, $E_{\phi_\epsilon}(F) \ge \epsilon\delta$.

Since $(\sigma_{\lambda_\epsilon})_X = \mu_X$ it also follows that

$$H(\phi_{\epsilon X}|\mu_X) = H((1-\epsilon)\mu_X + \epsilon\phi_X|\mu_X) = \int_{\Sigma_X} (1-\epsilon+\epsilon g)\log(1-\epsilon+\epsilon g) d\mu_X \, ,$$

where $g = \frac{d\phi_X}{d\mu_X} = \int_{\Sigma_Y} f d\mu_Y$. Expanding $[1+\epsilon(g-1)]\log[1+\epsilon(g-1)]$ up to second order in $\epsilon$ around $\epsilon = 0$ results with

$$\begin{aligned} H(\phi_{\epsilon X}|\mu_X) &= \epsilon \int_{\Sigma_X} (g-1) d\mu_X + \frac{1}{2}\epsilon^2 \int_{\Sigma_X} \frac{(g-1)^2}{(1+\xi(g-1))} d\mu_X \\ &\le \frac{1}{2}\epsilon^2 \int_{\Sigma_X} \frac{(g-1)^2}{(1-\epsilon|g-1|)} d\mu_X \, , \end{aligned}$$

where $\xi(x) \in [0,\epsilon]$. Hence, with $g$ bounded above by $1/a > 1$

$$J_X(\phi_\epsilon) = E_{\phi_\epsilon}(F)/H(\phi_{\epsilon X}|\mu_X) \ge \epsilon\delta/(\frac{1}{2}\epsilon^2/a^2(1-\epsilon/a)) \to \infty \text{ as } \epsilon \to 0,$$

while $H(\phi_\epsilon|\mu_X \times \mu_Y) < \infty$ for all $\epsilon > 0$, as required.

Paraphrasing the above argument, if $E_{\mu_Y}[G_Y] \ge 0$, then there exists $\sigma \in M_1(\Sigma)$ with $J_Y(\sigma)$ arbitrarily large and $H(\sigma|\mu_X \times \mu_Y) < \infty$. ∎

## References


Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215** 403-410.

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

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.

Arratia, R., Morris, P. and Waterman, M. S. (1988). Stochastic scrabble: large deviations for sequences with scores. *J. Appl. Probab.* **25** 106-119.

Arratia, R. and Waterman, M. S. (1985). Critical phenomena in sequence matching. *Ann. Probab.* **13** 1236-1249.

Arratia, R. and Waterman, M. S. (1989). The Erdös-Rényi strong law for pattern matching with a given proportion of mismatches. *Ann. Probab.* **17** 1152-1169.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, Wiley, New York.

Csiszár, I. and Körner, J. K. (1981). *Information theory: coding theorems for discrete memoryless systems.* Academic Press, New York.

Dembo, A., Cover, T. M. and Thomas, J. A. (1991). Information theoretic inequalities. *IEEE Trans. Inf. Theory* **37** 1501-1518.

Dembo, A. and Karlin, S. (1991a). Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables. *Ann. Probab.* **19** 1737-1755.

Dembo, A. and Karlin, S. (1991b). Strong limit theorems of empirical distributions for large segmental exceedances of partial sums of Markov variables. *Ann. Probab.* **19** 1755-1767.

Dembo, A. and Zeitouni, O. (1993). *Large Deviations Techniques and Applications*, Jones and Bartlett.

Dembo, A., Karlin, S. and Zeitouni, O. (1994). Limit distribution of maximal non-aligned two-sequence segmental score. To appear in Ann. Probab.

Deuschel, J. D. and Stroock, D. W. (1989). *Large Deviations*, Academic.

Erdös, P. and Rényi, A. (1970). On a new law of large numbers. *J. Analyse Math.* **22** 103-111.

Han, T. S. (1978). Nonnegative entropy measures of multivariate symmetric correlations. *Inform. Contr.* **36** 133-156.

Henikoff, S, and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA*, in press.

Karlin S. and Dembo A. (1992). Limit distributions of maximal segmental score among Markov dependent partial sums. *Advances in Applied Probability*, **24** 113-140.

Karlin, S. and Ost, F. (1988). Maximal length of common words among random letter sequences. *Ann. Probab.* **16** 535-563.

States, D. J., Gish, W. and Altschul, S. F. (1991). Improved sensitivity of nucleic acid database searches using application-specific scoring matrices. *Methods*, **3** 66-70.

Stormo, G. D. and Hartzell, G. W. (1989). Identifying protein-binding sites from unaligned DNA fragments. *Proc. Natl. Acad. Sci. USA*, **86** 1183-1187.