On the longest common increasing binary subsequence
Sur la plus longue sous-suite binaire croissante et commune

Christian Houdré[*], Jüri Lember[†], Heinrich Matzinger

July 24, 2006

Abstract

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two independent sequences of iid Bernoulli random variables with parameter 1/2. Let LCI_n be the length of the longest increasing sequence which is a subsequence of both finite sequences $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$. We prove that, as n goes to infinity, $n^{-1/2}(LCI_n - n/2)$ converges in law to a Brownian functional that we identify.

Résumé

Soient $X_1, X_2, \ldots$ et $Y_1, Y_2, \ldots$ deux suites mutuellement indépendantes de variables aléatoires de Bernoulli indépendantes, équidistribuées de paramètre 1/2. Soit LCI_n la longueur de la plus longue sous-suite croissante et commune aux deux sous-suites finies $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$. Nous démontrons que $n^{-1/2}(LCI_n - n/2)$ converge en loi vers une fonctionnelle brownienne que nous identifions.

Version française abrégée

Dans cette Note, nous nous intéressons au problème de déterminer l'ordre des fluctuations de la longueur de la plus longue sous-suite croissante qui est aussi commune à deux suites binaires indépendantes. Notre résultat principal est le suivant:

Théorème. Soient $X_1, X_2, \ldots, X_n, \ldots$ et $Y_1, Y_2, \ldots, Y_n, \ldots$ deux suites indépendantes de variables aléatoires de Bernoulli iid et de paramètre 1/2. Soit LCI_n le maximum des entiers $1 \leq k \leq n$ tels qu'il existe $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ et $1 \leq j_1 < j_2 < \ldots < j_k \leq n$ avec $X_{i_1} \leq X_{i_2} \leq \ldots \leq X_{i_k}$, $Y_{j_1} \leq Y_{j_2} \leq \ldots \leq Y_{j_k}$ et $X_{i_s} = Y_{j_s}$ pour tout $s = 1, \ldots, k$. Alors

$$\frac{LCI_n - n/2}{\sqrt{n}} \Longrightarrow \max_{t \in [0,1]} \left[\min_{i=1,2} \left(B^i(t) - \frac{1}{2} B^i(1) \right) \right],$$

où $B^1 = \left(B^1(t)\right)_{t\in[0,1]}$ et $B^2 = \left(B^2(t)\right)_{t\in[0,1]}$ sont deux mouvements browniens standards indépendants.

Les motivations et perspectives de notre travail sont à la fois les interactions profondes entre les problèmes de sous-suites et diverses branches des mathématiques comme la combinatoire algébrique, les matrices aléatoires, et les polynômes orthogonaux, mais aussi des considérations plus pratiques de bioinformatique. L'obtention de lois limites, souvent nouvelles, et dans le cadre d' alphabets finis, de modèles markoviens, ... est en effet une première étape vers le développement de méthodes quantitatives, telles que des tests statistiques, utilisables en séquencage d'ADN.

1 Introduction: The One Sequence Case

Longest increasing subsequence (LIS) problems have recently enjoyed renewed popularity. This stems mainly from the work of Baik, Deift and Johansson ([1]) who showed that the limiting law of the fluctuations of the longest increasing subsequence of a random permutation is the same as the law of the maximal eigenvalue

[*]All the authors would like to thank F. Götze, and the financial support of SFB 701 A3, at the University of Bielefeld, for supporting this research.

[†]Supported by the Estonian Science Foundation Grant nr. 5694

of certain random matrix models. This result has led to numerous recent advances, such as the work of Borodin ([3]) for colored random permutations or the case of finite alphabets random words by Its, Tracy and Widom ([8], [4], [5]) as well as Johansson ([6]). Related results with links to queueing theory were also obtained by Baryshnikov ([2]) and further connections with orthogonal polynomials, random matrices, growth models have since been further investigated by many authors.

A related important open problem, with direct consequences in computational biology (see, e.g. Waterman [9], [10]), is to find the order of the fluctuations of the length of the longest common subsequence (LCS) of two or more random sequences.

We obtain below the limiting distribution for the hybrid problem of the longest common and increasing subsequence (LCIS) of two random binary sequences. We start by presenting the one sequence case, where the results are known and obtained in the works just cited. Our approach might be worthwhile because of its simplicity and because it naturally leads and extends to the case of two (or more) sequences.

Let $X := (X_1, X_2, \ldots) \in \{0,1\}^{\mathbb{N}}$ be an infinite binary sequence. Let LI_n be the length of the longest increasing subsequence of $X_1, X_2, \ldots, X_n$, i.e. LI_n is the maximal $k \le n$ such that there exists an increasing sequence of natural numbers $1 \le i_1 < i_2 < \ldots < i_k \le n$ such that $X_{i_1} \le X_{i_2} \le \ldots \le X_{i_k}$. Let b_k be the number of ones in the finite sequence $X_1, X_2, \ldots, X_k$, in other words, let $b_0 := 0, \quad b_k := \sum_{i=1}^k X_i$, and let also a_k be the number of zeros in the sequence $X_1, X_2, \ldots, X_k$. Clearly, $a_k = k - b_k$. Next, for every $0 \le k \le n$, an increasing subsequence of $X_1, X_2, \ldots, X_n$ can be constructed by taking all the zeros up to (including) X_k, and then by taking all the ones between (and including) X_{k+1} and X_n. The number of zeros up to time k is equal to a_k, while the number of ones from X_k to X_n is equal to $b_n - b_k$. The maximum over $k = 0, \ldots, n$ of the length of all the subsequences obtained in this way is LI_n. In other words,

$$LI_n = \max_{k=0,\ldots,n} (a_k + (b_n - b_k)) = b_n + \max_{k=0,\ldots,n} (k - 2b_k).$$

Letting $Z_i := 1 - 2X_i$, it is then clear that $a_k - b_k = k - 2b_k = \sum_{i=1}^k Z_i$, and so setting $S_0 = 0, S_k = \sum_{i=1}^k Z_i$, $k \ge 1$, gives

$$LI_n = \frac{n}{2} - \frac{S_n}{2} + \max_{k=0,\ldots,n} S_k. \tag{1.1}$$

Using (1.1) and the reflection principle, one easily shows that $\mathbf{E}LI_n = n/2 + \sqrt{2n/\pi} + o(\sqrt{n})$ and that $\mathbf{Var}LI_n = 3n/4 - 2n/\pi + o(n)$. Moreover, Donsker's theorem and the continuous mapping theorem yield

$$\frac{LI_n - n/2}{\sqrt{n}} \Longrightarrow -\frac{B(1)}{2} + \max_{t\in[0,1]} B(t), \tag{1.2}$$

where "$\Longrightarrow$" stands for convergence in law. The above limiting law is well known and connected to a theorem of Pitman (see [7]). Its density can also be derived "by hand" and is given by: $16x^2 e^{-2x^2}/\sqrt{2\pi}, x > 0$.

2 The Two Sequence Case

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two binary sequences, and let

$$X^n := (X_1, \ldots, X_n), \quad Y^n := (Y_1, \ldots, Y_n).$$

Denote by LCI_n the length of the longest common increasing subsequence which is contained in both X^n and Y^n. In other words, LCI_n is the maximum over the ks that satisfy the following condition: there exist $1 \le i_1 < i_2 < \ldots < i_k \le n$ and $1 \le j_1 < j_2 < \ldots < j_k \le n$ such that

$$X_{i_1} \le X_{i_2} \le \ldots \le X_{i_k}, \quad Y_{j_1} \le Y_{j_2} \le \ldots \le Y_{j_k}$$

and $X_{i_s} = Y_{j_s}$ for all $s = 1, \ldots, k$. Let N_1 (resp. N_2) be the number of zeros in X^n (resp. in Y^n). Let T_k^1 denote the location of the k^{th} zero in the sequence $(X_1, X_2, \ldots)$, i.e. T_k^1 is defined recursively by the equations

$$T_0^1 = 0, \quad T_1^1 = \min\{t : X_t = 0\}, \quad T_{k+1}^1 = \min\{t > T_k^1 : X_t = 0\}.$$

In a similar way define T_k^2 to be the location of the k^{th} zero in the sequence $(Y_1, Y_2, \ldots)$. Let

$$g^1 : \{0, \ldots, N_1\} \to \mathbb{N} \text{ (resp.} g^2 : \{0, \ldots, N_2\} \to \mathbb{N})$$

be the maximum number of ones contained in any increasing subsequence of X^n(resp. of Y^n) which contains exactly k zeros. Hence, $g^1(k) = \sum_{i>T_k^1}^n X_i$, $k = 0,\ldots,N_1$, $g^2(k) = \sum_{i>T_k^2}^n Y_i$, $k = 0,\ldots,N_2$, and, in particular, $g^i(0) = n - N_i$. Thus, $g^1(k) + k$ (resp. $g^2(k) + k$) is the length of the longest increasing subsequence of X^n (resp. Y^n) that contains exactly k zeros and $\min_{i=1,2} g^i(k) + k$ is the length of the longest common increasing subsequence with exactly k zeros. Since a common subsequence of X and Y can contain at most $N_1 \wedge N_2$ zeros, we have the following useful relation:

$$LCI_n = \max_{k=0,\ldots,N_1\wedge N_2} \left[\min_{i=1,2}(g^i(k) + k)\right]. \tag{2.1}$$

Let $1 \le k \le N_i$. Between $k-1$ and k, the function g^i decreases by the number of ones located between T_{k-1}^i and T_k^i. This number is equal to $Z_k^i := T_k^i - T_{k-1}^i - 1$, $k = 1,\ldots,N_i$. Thus, for $i = 1,2$, it follows that

$$g^i(k) - g^i(k-1) = -Z_k^i, \quad k = 1,\ldots,N_i. \tag{2.2}$$

Moreover, recall that $g^i(0) = n - N_i$, and thus for any $k \ge 1$,

$$g^i(k) = n - N_i - \sum_{j=1}^k Z_j^i, \quad i = 1,2. \tag{2.3}$$

Assume now that the sequences $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are independent of each other. Let also the X_ks as well as the Y_ks be iid Bernoulli random variables with parameter $1/2$. In this case, $T_1^i, T_2^i, \ldots, T_k^i, \ldots$ are Pascal (negative binomial) random variables with respective parameters $1,2,\ldots,k,\ldots$ and $1/2$ and, as such each T_k^i is the sum of k iid geometric random variables with parameter $1/2$. Now, for $i = 1,2$, $Z_1^i + 1, Z_2^i + 1, Z_3^i + 1\ldots$ is the corresponding sequence of iid geometric random variables with parameter $1/2$. Hence $Z_1^i, Z_2^i, Z_3^i \ldots,$ is a sequence of iid random variables with $\mathbf{E}(Z_1^i) = 1$ and $\mathbf{Var} Z_1^i = 2$. Moreover the sequences $Z_1^1, Z_2^1, Z_3^1 \ldots$ and $Z_1^2, Z_2^2, Z_3^2 \ldots$ are also independent. We use the sequence $Z_1^i, Z_2^i, Z_3^i \ldots$ to approximate a standard Brownian motion. Let $k = 0,\ldots,n$ and $t = k/n$, and let $\hat{B}_n^i(t) := -\sum_{j=1}^{tn}(Z_j^i - 1)/\sqrt{2n}$. For $t \in (k/n, (k+1)/n)$, $k = 0,1,\ldots,n-1$, again define $\hat{B}_n^i(t)$ by linear interpolation. By (2.2) and (2.3), it thus follows that $g^i(k) + k = g^i(0) - \sum_{j=1}^k (Z_j^i - 1) = n - N_i + \sqrt{2n}\hat{B}_n^i(k/n)$. Hence, by (2.1)

$$LCI_n = \max_{0\le k\le N_1\wedge N_2} \left[\left(n - N_1 + \sqrt{2n}\hat{B}_n^1\left(\frac{k}{n}\right)\right) \wedge \left(n - N_2 + \sqrt{2n}\hat{B}_n^2\left(\frac{k}{n}\right)\right)\right]. \tag{2.4}$$

Note that

$$T_k^i = \sum_{j=1}^k (Z_j^i + 1) = \sum_{j=1}^k (Z_j^i - 1) + 2k = -\sqrt{2n}\hat{B}_n^i\left(\frac{k}{n}\right) + 2k. \tag{2.5}$$

Moreover, N_i is a binomial random variable with parameters n and $1/2$, and thus for n large it is highly concentrated around its mean $n/2$.

The next theorem states that this setting an analogue of (1.2) hold, i.e. $n^{-1/2}(LCI_n - n/2)$ converges in law to a Brownian functional. This implies that LI_n as well as the stochastically smaller random variable LCI_n have the same order of fluctuation. Intuitively, to better understand this result, note that it is very likely that, in both sequences, the first $n/2$ terms contains around $n/4$ zeroes, and the last $n/2$ terms contains around $n/4$ ones. When this is the case, the longest common subsequence is at least $n/2$. However, since by (1.2), LI_n is equivalent to $n/2$, the stochastically smaller random variable LCI_n should have the same property.

Theorem 2.1 *Let $X_1, X_2, \ldots, X_n, \ldots$ and $Y_1, Y_2, \ldots, Y_n, \ldots$ be two independent sequences of iid Bernoulli random variables with parameter $1/2$. Then*

$$\frac{LCI_n - n/2}{\sqrt{n}} \Longrightarrow \max_{t\in[0,1]} \left[\min_{i=1,2}\left(B^i(t) - \frac{1}{2}B^i(1)\right)\right], \tag{2.6}$$

where $B^1 = \left(B^1(t)\right)_{t\in[0,1]}$ and $B^2 = \left(B^2(t)\right)_{t\in[0,1]}$ are two independent standard Brownian motions.

Proof. By the self-similarity property of Brownian motion, to prove (2.6) it suffices to show that

$$\frac{LCI_n - n/2}{\sqrt{2n}} \Rightarrow \max_{t\in\left[0,\frac{1}{2}\right]} \left[\min_{i=1,2}\left(B^i(t) - \frac{1}{2}B^i\left(\frac{1}{2}\right)\right)\right]. \tag{2.7}$$

Next, for a_k, b_k reals,

$$|\max_{k=1,\dots,n}(a_k \wedge b_k) - \max_{k=1,\dots,n}((a_k+c)\wedge(b_k+d))| \le \max_{k=1,\dots,n}|(a_k\wedge b_k) - ((a_k+c)\wedge(b_k+d))| \le |c|\vee|d|. \tag{2.8}$$

By (2.4),

$$D_n := \frac{LCI_n - n/2}{\sqrt{2n}} = \max_{0\le k\le N_1\wedge N_2}\left[\left(\frac{n/2-N_1}{\sqrt{2n}} + \hat{B}_n^1\left(\frac{k}{n}\right)\right)\wedge\left(\frac{n/2-N_2}{\sqrt{2n}} + \hat{B}_n^2\left(\frac{k}{n}\right)\right)\right].$$

Let

$$\gamma_n^i := \frac{n/2-N_i}{\sqrt{2n}} + \frac{1}{2}\hat{B}_n^i\left(\frac{N_i}{n}\right), \quad i=1,2.$$

So

$$D_n = \max_{0\le k\le N_1\wedge N_2}\left[\left(\gamma_n^1 - \frac{1}{2}\hat{B}_n^1\left(\frac{N_1}{n}\right) + \hat{B}_n^1\left(\frac{k}{n}\right)\right)\wedge\left(\gamma_n^2 - \frac{1}{2}\hat{B}_n^2\left(\frac{N_2}{n}\right) + \hat{B}_n^2\left(\frac{k}{n}\right)\right)\right].$$

Let

$$U_n := \max_{0\le k\le N_1\wedge N_2}\left[\left(-\frac{1}{2}\hat{B}_n^1\left(\frac{N_1}{n}\right) + \hat{B}_n^1\left(\frac{k}{n}\right)\right)\wedge\left(-\frac{1}{2}\hat{B}_n^2\left(\frac{N_2}{n}\right) + \hat{B}_n^2\left(\frac{k}{n}\right)\right)\right].$$

By (2.8),

$$|D_n - U_n| \le |\gamma_n^1|\vee|\gamma_n^2|. \tag{2.9}$$

Let

$$V_n := \max_{0\le k\le N_1\wedge N_2}\left[\left(-\frac{1}{2}\hat{B}_n^1\left(\frac{1}{2}\right) + \hat{B}_n^1\left(\frac{k}{n}\right)\right)\wedge\left(-\frac{1}{2}\hat{B}_n^2\left(\frac{1}{2}\right) + \hat{B}_n^2\left(\frac{k}{n}\right)\right)\right].$$

By (2.8),

$$|U_n - V_n| \le \frac{1}{2}\left|\hat{B}_n^1\left(\frac{1}{2}\right) - \hat{B}_n^1\left(\frac{N_1}{n}\right)\right| \vee \frac{1}{2}\left|\hat{B}_n^2\left(\frac{1}{2}\right) - \hat{B}_n^2\left(\frac{N_2}{n}\right)\right|. \tag{2.10}$$

Let

$$X_n := \max_{0\le t\le 1/2}\left[\left(-\frac{1}{2}\hat{B}_n^1\left(\frac{1}{2}\right) + \hat{B}_n^1(t)\right)\wedge\left(-\frac{1}{2}\hat{B}_n^2\left(\frac{1}{2}\right) + \hat{B}_n^2(t)\right)\right].$$

Hence,

$$V_n - X_n \le \max_{t\in\left[\frac{1}{2},\frac{N_1}{n}\right]}\left(\hat{B}_n^1(t) - \hat{B}_n^1\left(\frac{1}{2}\right)\right) \vee \max_{t\in\left[\frac{1}{2},\frac{N_2}{n}\right]}\left(\hat{B}_n^2(t) - \hat{B}_n^2\left(\frac{1}{2}\right)\right). \tag{2.11}$$

In the following, let $i = 1, 2$ be fixed, and let us skip it from the notation. By the very definition of $\hat{B}_n$,

$$\max_{t\in\left[\frac{1}{2},\frac{N}{n}\right]}\left(\hat{B}_n(t) - \hat{B}_n\left(\frac{1}{2}\right)\right) = \max_{k=\lceil n/2\rceil,\dots,N}\left(\hat{B}_n\left(\frac{k}{n}\right) - \hat{B}_n\left(\frac{1}{2}\right)\right)\vee 0.$$

Let $m = \lceil n/2\rceil$, where $\lceil\cdot\rceil$ is the usual ceiling (or greatest integer) function. Then, $\hat{B}_n(k/n) - \hat{B}_n(1/2) = \sum_{j=m}^k \xi_j/\sqrt{2n}$, where

$$\xi_m = \sqrt{2n}\left(\hat{B}_n\left(\frac{m}{n}\right) - \hat{B}_n\left(\frac{1}{2}\right)\right), \; \xi_{m+1} = Z_{m+1}-1, \; \xi_{m+2} = Z_{m+2}-1, \dots, \xi_k = Z_k - 1.$$

Clearly, $\xi_m = 0,$ if n is even, and $\xi_m = \frac{1}{2}(Z_m^i - 1)$ otherwise. Let $C_n := \left\{ \left|\frac{N}{n} - \frac{1}{2}\right| \le \frac{\ln n}{\sqrt{n}} \right\}$, then

$$\begin{aligned}
\mathbf{P}\left(\max_{t\in\left[\frac{1}{2},\frac{N}{n}\right]} \left| \hat{B}_n(t) - \hat{B}_n\left(\frac{1}{2}\right) \right| > \varepsilon \right) &= \mathbf{P}\left(\max_{k=m,\dots,N} \frac{1}{\sqrt{2n}} \left| \sum_{j=m}^{k} \xi_j \right| > \varepsilon \right) \\
&\le \mathbf{P}\left(\max_{k=m,\dots,n/2+\sqrt{n}\ln n} \frac{1}{\sqrt{n}} \left| \sum_{j=m}^{k} \xi_j \right| > \varepsilon \right) + \mathbf{P}(C_n^c) \\
&\le \frac{\ln n}{\varepsilon^2\sqrt{n}} + \mathbf{P}(C_n^c),
\end{aligned}$$

by Kolmogorov'ss inequality. Next, $\mathbf{P}(C_n^c) \to 0$. Indeed, $N \sim \mathrm{Bin}(n, 1/2)$ and so,

$$\mathbf{P}(C_n^c) = \mathbf{P}(|N - n/2| > \sqrt{n}\log n) \le 2e^{-2n(\log n)^2/n} = 2e^{-2(\log n)^2}.$$

Thus,

$$\max_{t\in\left[\frac{1}{2},\frac{N}{n}\right]} \left(\hat{B}_n(t) - \hat{B}_n\left(\frac{1}{2}\right) \right) \xrightarrow{P} 0, \tag{2.12}$$

implying that $V_n - X_n \xrightarrow{P} 0$. Now,

$$X_n - V_n \le \max_{t\in\left[\frac{N_1}{n},\frac{1}{2}\right]} \left(\hat{B}_n^1(t) - \hat{B}_n^1\left(\frac{N_1}{n}\right) \right) \vee \max_{t\in\left[\frac{N_2}{n},\frac{1}{2}\right]} \left(\hat{B}_n^2(t) - \hat{B}_n^2\left(\frac{N_2}{n}\right) \right). \tag{2.13}$$

To prove that

$$\max_{t\in\left[\frac{N}{n},\frac{1}{2}\right]} \left(\hat{B}_n(t) - \hat{B}_n\left(\frac{N}{n}\right) \right) \xrightarrow{P} 0, \tag{2.14}$$

we use similar arguments, since

$$\begin{aligned}
\mathbf{P}\left(\max_{t\in\left[\frac{N}{n},\frac{1}{2}\right]} \left| \hat{B}_n(t) - \hat{B}_n\left(\frac{N}{n}\right) \right| > \varepsilon \right) &= \mathbf{P}\left(\max_{k=N,\dots,m} \frac{1}{\sqrt{2n}} \left| \sum_{j=N}^{k} \xi_j \right| > \varepsilon \right) \\
&\le \mathbf{P}\left(\max_{k=n/2-\sqrt{n}\ln n,\dots,m} \frac{1}{\sqrt{2n}} \left| \sum_{j=n/2-\sqrt{n}\ln n}^{k} \xi_j \right| > \varepsilon \right) \\
&\qquad + \mathbf{P}(C_n^c) \longrightarrow 0.
\end{aligned}$$

Hence, $X_n - V_n \xrightarrow{P} 0$, and so $|X_n - V_n| \xrightarrow{P} 0$. Together, the convergence results (2.12) and (2.14) imply that $|U_n - V_n| \xrightarrow{P} 0$. Let us next prove that $\gamma_n^i \xrightarrow{P} 0$. Again, we skip i from the notation. From (2.5),

$$\frac{n/2 - N}{\sqrt{2n}} = \frac{n - T_N}{2\sqrt{2n}} - \frac{1}{2}\hat{B}_n\left(\frac{N}{n}\right),$$

and

$$\gamma_n = \frac{n/2 - N}{\sqrt{2n}} + \frac{1}{2}\hat{B}_n\left(\frac{N}{n}\right) = \frac{n - T_N}{2\sqrt{2n}}.$$

Now, T_N is the location of the last zero in $X_1, \dots, X_n$, and so $\mathbf{P}(n - T_N = j) = 2^{-j+1}$, if $j = 0, \dots, n-1$ while $\mathbf{P}(n - T_N = n) = 2^{-n}$. Hence, for any $\varepsilon > 0$,

$$\mathbf{P}(|n - T_N| > 2\varepsilon\sqrt{2n}) = \mathbf{P}(n - T_N > 2\varepsilon\sqrt{2n}) \le \left(\frac{1}{2}\right)^{2\varepsilon\sqrt{2n}} \to 0.$$

The convergence of $\gamma_n \xrightarrow{P} 0$ follows. Hence, $|D_n - U_n| \xrightarrow{P} 0$, $|U_n - V_n| \xrightarrow{P} 0$ and $|X_n - V_n| \xrightarrow{P} 0$ and so

$$|D_n - X_n| \xrightarrow{P} 0. \tag{2.15}$$

Let Y^i, $i = 1, 2$ be a $C[0,1]$-valued random element so that

$$Y_n^i(t) := -\frac{1}{2}\hat{B}_n^i\left(\frac{1}{2}\right) + \hat{B}_n^i(t), \quad i = 1, 2.$$

Since $\hat{B}_n^i \Rightarrow B^i$, it follows that $Y_n^i \Rightarrow B^i - 2^{-1}B^i(1/2)$, $i = 1, 2$. Let $Y_n := (Y_n^1, Y_n^2)$. Then Y_n is a $C[0,1] \times C[0,1]$-valued random element. Since Y_n^1 and Y_n^2 as well as B^1 and B^2 are independent, $Y_n \Rightarrow (B^1, B^2)$. Appealing twice to the continuous mapping theorem shows that

$$X_n = \max_{t\in[0,\frac{1}{2}]} \left(Y_n^1(t) \wedge Y_n^2(t)\right) \Rightarrow \max_{t\in[0,\frac{1}{2}]} \left[\left(B^1(t) - \frac{1}{2}B^1\left(\frac{1}{2}\right)\right) \wedge \left(B^2(t) - \frac{1}{2}B^2\left(\frac{1}{2}\right)\right)\right].$$

By (2.15), D_n converges in distribution to the same limit. ■

It would be interesting to find a more explicit representation for the law of the limiting distribution obtained in the above theorem, in other words for the law of

$$\max_{t\in[0,1]} \frac{1}{\sqrt{2}}\left[B^1(t) - \frac{1}{2}B^1(1) - \left|B^2(t) - \frac{1}{2}B^2(1)\right|\right].$$

Note also that the proof of the above theorem can be easily extended to an arbitray, but fixed, number k of random sequences leading to the functional $\max_{t\in[0,1]}\left[\min_{i=1,\dots,k}\left(B^i(t) - B^i(1)/2\right)\right]$.

References

[1] J. Baik, P. Deift, K. Johansson, On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** (1999) 1119–1178.

[2] Yu. Baryshnikov, GUEs and queues. *Probab. Theory Related Fields* **119** (2001) 256–274.

[3] A. Borodin, Longest increasing subsequences of random colored permutations. *Electron. J. Combin.*, **6** (1999), Research Paper 13, 12 pp.

[4] A. Its, C. Tracy, H. Widom, Random words, Toeplitz determinants, and integrable systems. I. Random matrix models and their applications, 245–258, *Math. Sci. Res. Inst. Publ.* **40**, Cambridge Univ. Press, Cambridge, 2001.

[5] A. Its, C. Tracy, H. Widom, Random words, Toeplitz determinants and integrable systems. II. Advances in nonlinear mathematics and science. *Phys. D*, **152/153** (2001) 199–224.

[6] K. Johansson, Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math.* **153** (2001) 259–296.

[7] D. Revuz, M. Yor, Continuous martingales and Brownian motion. Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293. Springer-Verlag, Berlin, 1999.

[8] C. Tracy, H. Widom, On the distributions of the lengths of the longest monotone subsequences in random words. *Probab. Theory Related Fields* **119** (2001) 350–380.

[9] M. Waterman, Introduction to Computational Biology Chapman & Hall 1995.

[10] H. Zhang, Alignment of BLAST high-scoring segment pairs based on the longest increasing subsequence algorithm. *Bioinformatics* **19** (2003) 1391–1396.

C.H: School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332, USA
houdre@math.gatech.edu

J.L: Institute of Mathematical Statistics
Tartu University
Liivi 2-513 50409, Tartu, Estonia
jyril@ut.ee

H.M: School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332, USA
matzing@math.gatech.edu