ON THE LONGEST COMMON INCREASING BINARY SUBSEQUENCE

C. Houdré[*], J. Lember[†], H. Matzinger

December 5, 2005

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. We give an explicit representation of the limit law in terms of the maximum of two Brownian motions. As a warm up, we treat the one sequence case and find (when properly centered and normalized) the limiting law of LI_n, the length of the longest increasing subsequence of the finite sequence $X_1, \ldots, X_n$.

1 Introduction: The One Sequence Case

Longest increasing subsequence problems have enjoyed a lot of popularity in recent years stemming mainly from the work of Baik, Deift and Johansson ([1]) who obtained the limiting law for the case of random permutations while Borodin ([3]) obtained it for colored random permutations. Further limiting distributions were obtained for finite alphabets random words by Its, Tracy and Widom ([8], [4], [5]) as well as Johansson ([6]). Related results were also obtained by Baryshnikov ([2]).

We obtain below the limiting distribution for the hybrid problem of the longest common and increasing subsequence 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. It is nevertheless our belief that our approach is worthwhile because of its simplicity and because it will naturally lead and extend to two sequences, a case which at the present has not been obtained by other methods.

[*]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

1.1 Combinatorics

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 $k \leq n$ such that there exists an increasing sequence of natural numbers $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ such that $X_{i_1} \leq X_{i_2} \leq \ldots \leq X_{i_k}$. Let b_k be the number of ones in the finite sequence $X_1, X_2, \ldots, X_k$, in other words,

$$b_0 := 0, \quad b_k := \sum_{i=1}^{k} X_i.$$

Let a_k be the number of zeros in the sequence $X_1, X_2, \ldots, X_k$. Thus, $a_k = k - b_k$.
For every $0 \leq k \leq n$, 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 excluding) X_k and X_n. The number of zeros up to time k is equal to a_k. 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 subsequence obtained in this way is LI_n. In other words,

$$\begin{aligned} LI_n &= \max_{k=0,\ldots,n} \left(a_k + (b_n - b_k)\right) \\ &= b_n + \max_{k=0,\ldots,n} \left(k - 2b_k\right). \end{aligned}$$

Now let

$$Z_i := \begin{cases} 1 & \text{if } X_i = 0, \\ -1 & \text{if } X_i = 1, \end{cases}$$

i.e., let $Z_i = 1 - 2X_i$. It is then clear that $a_k - b_k = k - 2b_k = \sum_{i=1}^{k} Z_i$, and so setting $S_0 = 0, S_k = \sum_{i=1}^{k} Z_i,\ k \geq 1$, gives

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

1.2 Limiting Distribution

If the X_is are iid Bernoulli random variables with parameter $1/2$, then $Z_1, Z_2, \ldots$ are also iid random variables with $\mathbf{P}(Z_i = 1) = \mathbf{P}(Z_i = -1) = 1/2$. Hence, very classically and invoking the reflection principle, we get

$$\mathbf{E}LI_n = \frac{n}{2} + \mathbf{E}|S_n| - \frac{1}{2}\left(1 - \mathbf{P}(S_n = 0)\right),$$

and thus $\lim_{n\to\infty} \mathbf{E}LI_n/n = 1/2$. In fact, since as well known, $\lim_{n\to+\infty} \mathbf{E}|S_n|/\sqrt{n} = \sqrt{2/\pi}$, we have:

$$\mathbf{E}LI_n = \frac{n}{2} + \sqrt{\frac{2n}{\pi}} + o(\sqrt{n}). \tag{1.2}$$

Still invoking the reflection principle, for any $\varepsilon > 0$:

$$\sum_{n=1}^{+\infty} \mathbf{P}\left(\max_{k=0,\dots,n} S_k \geq n\varepsilon\right) \leq \sum_{n=1}^{+\infty} \frac{\mathbf{E}S_n^4}{n^4\varepsilon^4} = \sum_{n=1}^{+\infty} \frac{3n^2-2n}{n^4\varepsilon^4} < +\infty.$$

Combining this last fact with Borel's strong law imply that, with probability one, $\lim_{n\to\infty} LI_n/n = 1/2$.

Again, the reflection principle implies that

$$\mathbf{E}(\max_{k=0,\dots,n} S_k)^2 = \mathbf{E}S_n^2 - \mathbf{E}|S_n| + \frac{1}{2}\left(1 - \mathbf{P}(S_n = 0)\right)$$

and so as $n \to \infty$, and with the help of (1.2) we obtain that

$$\mathbf{E}\left(LI_n - \mathbf{E}LI_n\right)^2 = \frac{3n}{4} - \frac{2n}{\pi} + o(n). \tag{1.3}$$

Next, and still very classically, it is well known that by rescaling the simple symmetric random walk $S := (S_k)_{k\geq 0}$ one approximately obtains a standard Brownian motion B, i.e., $B = (B(t))_{t\in[0,1]}$ is a continuous version of Brownian motion with $B(0) = 0$, and $\mathbf{Var}B(t) = t$. For, $k = 0, \dots, n$, and $t = k/n$ let

$$\hat{B}_n(t) := \frac{S_k}{\sqrt{n}},$$

while for $t \in (k/n, (k+1)/n)$, $k = 1, \dots, n-1$, $\hat{B}_n(t)$ is defined by linear interpolation. Thus, $\hat{B}_n$ approximates B by using $Z_1, \dots, Z_n$. The equality (1.1) now yields

$$\begin{aligned} LI_n &= \frac{n}{2} - \frac{\sqrt{n}\hat{B}_n(1)}{2} + \max_{k=0,\dots,n} \sqrt{n}\hat{B}_n\left(\frac{k}{n}\right) \\ &= \frac{n}{2} + \sqrt{n}\left(-\frac{\hat{B}_n(1)}{2} + \max_{t\in[0,1]} \hat{B}_n(t)\right), \end{aligned} \tag{1.4}$$

since $\hat{B}_n(t)$ is linear between the points k/n and since then the above maximum can as well be taken over all $t \in [0,1]$. Now, since the process $\hat{B}_n$ converges to a standard Brownian motion, for large n, LI_n is "approximately equal to" $\frac{n}{2} + \sqrt{n}\left(-\frac{B(1)}{2} + \max_{t\in[0,1]} B(t)\right)$. A more precise formulation of this fact is given by:

Proposition 1.1 *Let $X_1, X_2, \dots, X_n, \dots,$ be iid Bernoulli random variables with parameter $1/2$, then*

$$\frac{LI_n - n/2}{\sqrt{n}} \Longrightarrow -\frac{B(1)}{2} + \max_{t\in[0,1]} B(t), \tag{1.5}$$

where "$\Longrightarrow$" stands for convergence in law.

Proof. By (1.4),

$$\frac{LI_n - n/2}{\sqrt{n}} = -\frac{\hat{B}_n(1)}{2} + \max_{t\in[0,1]} \hat{B}_n(t). \tag{1.6}$$

By Donsker's theorem, $\hat{B}_n \Rightarrow B$ in the space $C[0,1]$ endowed with the supremum norm. Since

$$g : C[0,1] \to \mathbb{R}, \quad g(x) = \sup_{t\in[0,1]} x(t) - \frac{x(1)}{2},$$

is a continuous function, by the continuous mapping theorem,

$$\frac{LI_n - n/2}{\sqrt{n}} = g(\hat{B}_n) \Rightarrow g(B) = -\frac{B(1)}{2} + \max_{t\in[0,1]} B(t).$$

■

1.3 Density of the Limiting Distribution

Although, we could obtain this density via a well known theorem of Pitman (see [7]), below, we derive it "by hand" both for the sake of simplicity and completeness and also since it was done in this way before we became aware of Pitman's result.

Proposition 1.2 *Let B be a standard Brownian motion, and let*

$$W := \max_{t\in[0,1]} B(t) - \frac{B(1)}{2}.$$

Then W is a positive random variable with density

$$f_W(w) = \frac{16w^2}{\sqrt{2\pi}} e^{-2w^2}, \quad w > 0.$$

Proof. Clearly, $W \geq \max(-B(1)/2, B(1)/2) \geq 0$. Let $Z := \max_{t\in[0,1]} B(t)$ and let

$$F(z,b) := P\left(Z \geq z, B(1) \leq b\right).$$

Let $\bar{\Phi}$ denote the survival function of the standard normal random variable:

$$\bar{\Phi}(s) := \int_s^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Let $z > 0$ and $b < z$. By the reflection principle,

$$P\left(Z \geq z, B(1) \leq b\right) = P\left(Z \geq z, B(1) \geq z + (z-b)\right) = P\left(B(1) \geq z + (z-b)\right).$$

Since $B(1)$ is standard normal, the right hand side of the last equation is equal to $\bar{\Phi}(2z-b)$, and so

$$F(z,b) = \bar{\Phi}(2z-b).$$

The joint density of Z and $B(1)$, denoted by f, can now be computed

$$f = -\frac{\partial^2 F}{\partial z \partial b}.$$

Since

$$-\frac{\partial F}{\partial z} = -2\bar{\Phi}'(2z-b) = \frac{2}{\sqrt{2\pi}} e^{-(2z-b)^2/2}$$

and

$$\frac{\partial e^{-(2z-b)^2/2}}{\partial b} = (2z-b)e^{-(2z-b)^2/2},$$

we have

$$f(z,b) = \frac{4z-2b}{\sqrt{2\pi}} e^{-(2z-b)^2/2} I_{\{z\geq 0\}} I_{\{b\leq z\}}.$$

Let $W := Z - B(1)/2$. The density of W can now be computed via the transformation $(z,b) \mapsto (z, z-b/2)$. The density of $(Z, Z-B(1)/2)$, denoted by $g(z,w)$, is

$$g(z,w) = \frac{8w}{\sqrt{2\pi}} e^{-2w^2} I_{\{2w\geq z\geq 0\}}.$$

Integrating over z gives the density of W:

$$f_W(w) = \int g(z,w)dz = \int_0^{2w} \frac{8w}{\sqrt{2\pi}} e^{-2w^2} dz = \frac{16w^2}{\sqrt{2\pi}} e^{-2w^2}.$$

■

The results of this section show that

$$\frac{LI_n - \mathbf{E}LI_n}{\sqrt{\mathbf{Var}LI_n}} \Longrightarrow L,$$

where L has density given by

$$\sqrt{\frac{3\pi-8}{2}}\frac{8}{\pi}\left(x\sqrt{\frac{3\pi-8}{4\pi}} + \sqrt{\frac{2}{\pi}}\right)^2 e^{-2\left(x\sqrt{\frac{3\pi-8}{4\pi}}+\sqrt{\frac{2}{\pi}}\right)^2}, \quad x > -\sqrt{\frac{8}{3\pi-8}}.$$

2 The Two Sequence Case

2.1 Combinatorics

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 ks that satisfy the following condition: There exist $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ and $1 \leq j_1 < j_2 < \ldots < j_k \leq n$ such that

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

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,

$$g^1(k) = \sum_{i > T_k^1}^{n} X_i, \quad k = 0, \ldots, N_1, \quad g^2(k) = \sum_{i > T_k^2}^{n} Y_i, \quad k = 0, \ldots, N_2,$$

and, in particular, $g^i(0) = n - N_i$. Let $D_1 \in \mathbb{N} \times \mathbb{N}$ (resp. $D_2 \in \mathbb{N} \times \mathbb{N}$) denote the surface under the curve g^1 (resp. g^2). Thus, $(x, y) \in D_i$ if and only if $x = 0, \ldots, N_i$ and $y \leq g^i(x)$, $i = 1, 2$.
Note that for $i = 1, 2$, we have that D_i, N_i, g^i depend on n. However, in order not to overburden the notation, n is ommited.
Let us next show how LCI_n can be written as the solution of an optimization problem involving g^1 and g^2.

Lemma 2.1

$$LCI_n = \max_{(x,y) = D_1 \cap D_2} x + y \tag{2.1}$$

$$LCI_n = \max_{k=0,\ldots,N_1 \wedge N_2} \left[\min_{i=1,2} (g^i(k) + k) \right]. \tag{2.2}$$

Proof. Let us start by proving (2.1). Assume that we have found an increasing subsequence with exactly x zeros and y ones. Then the total length of this subsequence is $x + y$. This increasing sequence is a subsequence of both X^n and Y^n if and only if $(x, y) \in D_1 \cap D_2$. Indeed, there are no more than N_1 zeros in X^n. Thus if the increasing sequence with x zeros and y ones is a subsequence of X^n, then $x \leq N_1$. Furthermore, the maximum number of ones in an increasing subsequence of X^n with x zeros in it, is the total number of ones between the x-th zero and n. This, by definition, is equal to $g_1(x)$. Thus, if the increasing sequence with x zeros and y ones is a subsequence of X^n, then

$y \leq g_1(x)$ and so $(x,y) \in D_1$. Similarly, if the increasing sequence with x zeros and y ones is a subsequence of Y^n, then $(x,y) \in D_2$. This implies that if an increasing sequence with x zeros and y ones is a subsequence of X^n and Y^n, then $(x,y) \in D_1 \cap D_2$.
On the other hand, if $(x,y) \in D_1 \cap D_2$, then it is possible to find a common subsequence of X^n and Y^n with x zeroes and y ones. Hence (2.1) follows.

Let us now prove (2.2). To do so, let $\partial(D_1 \cap D_2)$ denote the boundary of the set $D_1 \cap D_2$, i.e.

$$\partial(D_1 \cap D_2) := \left\{ (x, \min_{i=1,2} g^i(x)) \in \mathbb{N} \times \mathbb{N} : x = 0, \ldots, N_1 \wedge N_2] \right\}.$$

The maximum of the function $(x,y) \mapsto x+y$ on the set $D_1 \cap D_2$ can only be attained on the boundary of $D_1 \cap D_2$. Hence

$$LCI_n = \max_{(x,y) \in \partial(D_1 \cap D_2)} x + y = \max_{x=0,\ldots,N_1 \wedge N_2} \left(x + \min_{i=1,2} g^i(x) \right).$$

■

Let $1 \leq k \leq N_i$. Between $k-1$ and k, the function g^i decreases by the number of ones located between T^i_{k-1} and T^i_k. This number is equal to

$$Z^i_k := T^i_k - T^i_{k-1} - 1, \quad k = 1, \ldots, N_i.$$

Thus, for $i = 1, 2$, it follows that

$$g^i(k) - g^i(k-1) = -Z^i_k, \quad k = 1, \ldots, N_i. \tag{2.3}$$

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

$$g^i(k) = n - N_i - \sum_{j=1}^{k} Z^i_j, \tag{2.4}$$

$i = 1, 2$.

2.2 Limiting Distribution

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 variables with parameter 1/2. In this case, $T^i_1, T^i_2, \ldots, T^i_k, \ldots$ are Pascal (negative binomial) random variables with respective parameters $1, 2, \ldots, k, \ldots$ and 1/2 and, as such each T^i_k is the sum of k iid geometric random variables with parameter 1/2. Now, for $i = 1, 2$

$$Z^i_1 + 1, Z^i_2 + 1, Z^i_3 + 1 \ldots$$

is the corresponding sequence of iid geometric random variables with parameter 1/2. Hence $Z^i_1, Z^i_2, Z^i_3 \ldots,$ is a sequence of iid random variables with $\mathbf{E}(Z^i_1) = 1$ and $\mathbf{Var} Z^i_1 = 2$.

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 define

$$\hat{B}_n^i(t) := -\frac{\sum_{j=1}^{tn}(Z_j^i - 1)}{\sqrt{2n}}.$$

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 (2.3) and (2.4), it thus follows that

$$g^i(k) + k = g^i(0) - \sum_{j=1}^{k}(Z_j^i - 1) = n - N_i + \sqrt{2n}\hat{B}_n^i\left(\frac{k}{n}\right).$$

Hence, by (2.2)

$$LCI_n = \max_{0 \le k \le N_1 \wedge N_2}\left[\left(n - N_1 + \sqrt{2n}\hat{B}_n^1\left(\frac{k}{n}\right)\right) \wedge \left(n - N_2 + \sqrt{2n}\hat{B}_n^2\left(\frac{k}{n}\right)\right)\right]. \quad (2.5)$$

Note that

$$T_k^i = \sum_{j=1}^{k}(Z_j^i + 1) = \sum_{j=1}^{k}(Z_j^i - 1) + 2k = -\sqrt{2n}\hat{B}_n^i\left(\frac{k}{n}\right) + 2k. \quad (2.6)$$

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 $n/2$. These notations and facts will give us:

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*

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

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

Proof. The selfsimilarity property of Brownian motion $B = B(t)_{t\in[0,1]}$, implies that

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

So, to prove (2.7), it suffices to show that

$$\frac{LCI_n - n/2}{\sqrt{2n}} \Rightarrow \max_{t\in\left[0,\frac{1}{2}\right]}\left[\min_{i=1,2}\left(B^i(t) - \frac{1}{2}B^i\left(\frac{1}{2}\right)\right)\right]. \quad (2.8)$$

Let a, b, c, d be reals. Then

$$|(a \wedge b) - (a + c) \wedge (b + d)| \leq |c| \vee |d|.$$

Hence, for a_k, b_k reals,

$$\begin{aligned} | \max_{k=1,\ldots,n} (a_k \wedge b_k) - \max_{k=1,\ldots,n} ((a_k + c) \wedge (b_k + d)) \, | &\leq \max_{k=1,\ldots,n} |(a_k \wedge b_k) - ((a_k + c) \wedge (b_k + d))| \\ &\leq |c| \vee |d|. \end{aligned} \tag{2.9}$$

By (2.5),

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

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

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

Let

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

By (2.9),

$$|D_n - U_n| \leq |\gamma_n^1| \vee |\gamma_n^2|. \tag{2.10}$$

Let

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

By (2.9),

$$|U_n - V_n| \leq \frac{1}{2} \left| \hat{B}_n^1 \left(\frac{1}{2} \right) - \hat{B}_n^1 \left(\frac{N_1}{n} \right) \right| \vee \frac{1}{2} \left| \hat{B}_n^2 \left(\frac{1}{2} \right) - \hat{B}_n^2 \left(\frac{N_2}{n} \right) \right|. \tag{2.11}$$

Let

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

Hence,

$$V_n - X_n \leq \max_{t \in \left[\frac{1}{2}, \frac{N_1}{n} \right]} \left(\hat{B}_n^1(t) - \hat{B}_n^1 \left(\frac{1}{2} \right) \right) \vee \max_{t \in \left[\frac{1}{2}, \frac{N_2}{n} \right]} \left(\hat{B}_n^2(t) - \hat{B}_n^2 \left(\frac{1}{2} \right) \right). \tag{2.12}$$

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

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

Let $m = \lceil n/2 \rceil$, where $\lceil \cdot \rceil$ is the usual ceiling (or greatest interger) function, then

$$\hat{B}_n\left(\frac{k}{n}\right) - \hat{B}_n\left(\frac{1}{2}\right) = \frac{1}{\sqrt{2n}} \sum_{j=m}^{k} \xi_j,$$

where

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

Clearly,

$$\xi_m = \begin{cases} 0, & \text{if } n \text{ is even,} \\ \frac{1}{2}(Z_m^i - 1) & \text{otherwise.} \end{cases}$$

Let

$$C_n := \left\{ \left| \frac{N}{n} - \frac{1}{2} \right| \leq \frac{\ln n}{\sqrt{n}} \right\}.$$

Note that

$$\left\{ \max_{k=m,\ldots,N} \frac{1}{\sqrt{2n}} \left| \sum_{j=m}^{k} \xi_j \right| > \varepsilon \right\} \subset \left\{ \max_{k=m,\ldots,n/2+\sqrt{n}\ln n} \frac{1}{\sqrt{2n}} \left| \sum_{j=m}^{k} \xi_j \right| > \varepsilon \right\} \cup C_n^c.$$

By Kolmogorov's inequality,

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

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

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

Thus,

$$\max_{t\in\left[\frac{1}{2},\frac{N}{n}\right]}\left(\hat{B}_n(t)-\hat{B}_n\left(\frac{1}{2}\right)\right)\xrightarrow{P}0, \tag{2.13}$$

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

Now

$$X_n - V_n \le \max_{t\in\left[\frac{N_1}{n},\frac{1}{2}\right]}\left(\hat{B}^1_n(t)-\hat{B}^1_n\left(\frac{N_1}{n}\right)\right)\vee\max_{t\in\left[\frac{N_2}{n},\frac{1}{2}\right]}\left(\hat{B}^2_n(t)-\hat{B}^2_n\left(\frac{N_2}{n}\right)\right). \tag{2.14}$$

To prove that

$$\max_{t\in\left[\frac{N}{n},\frac{1}{2}\right]}\left(\hat{B}_n(t)-\hat{B}_n\left(\frac{N}{n}\right)\right)\xrightarrow{P}0, \tag{2.15}$$

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 (2.13) and (2.15) imply that

$$\mathbf{P}\left(\left|\hat{B}_n\left(\frac{1}{2}\right)-\hat{B}_n\left(\frac{N}{n}\right)\right|>\varepsilon\right)\le\mathbf{P}\left(\max_{t\in\left[\frac{1}{2},\frac{N}{n}\right]}\left|\hat{B}_n(t)-\hat{B}_n\left(\frac{1}{2}\right)\right|>\varepsilon\right)\to 0,$$

i.e. $|U_n - V_n| \xrightarrow{P} 0$.

Let us next prove that $\gamma_n^i \xrightarrow{P} 0$. Again, we skip i from the notation. By (2.6),

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

Hence,

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

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 $X_1, \ldots, X_n$, and so $\mathbf{P}(n - T_N = j) = 2^{-j+1}$, if $j = 0, \ldots, n-1$ 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

$$|D_n - X_n| \xrightarrow{P} 0. \tag{2.16}$$

Let Y^i, $i = 1, 2$ be $C[0,1]$-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 holds that

$$Y_n^i \Rightarrow B^i - \frac{1}{2}B^i\left(\frac{1}{2}\right), \quad i = 1, 2.$$

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,

$$Y_n \Rightarrow (B^1, B^2).$$

Let, for every $(y_1, y_2) \in C[0,1] \times C[0,1]$, $\|(y_1, y_2)\| := \|y_1\| + \|y_2\|$. This metric generates the product σ-field on $C[0,1] \times C[0,1]$. The mapping

$$f : C[0,1] \times C[0,1] \to C[0,1], \quad f(y_1, y_2) = y_1 \wedge y_2$$

is continuous. By the continuous mapping theorem,

$$Y_n^1 \wedge Y_n^2 \Rightarrow \left(B^1 - \frac{1}{2}B^1\left(\frac{1}{2}\right)\right) \wedge \left(B^2 - \frac{1}{2}B^2\left(\frac{1}{2}\right)\right).$$

Again, by the continuous mapping theorem,

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

By (2.16), 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

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

References

[1] Baik, Jinho; Deift, Percy; Johansson, Kurt. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** (1999), no. 4, 1119–1178.

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

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

[4] Its, Alexander R.; Tracy, Craig A.; Widom, Harold. 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] Its, Alexander R.; Tracy, Craig A.; Widom, Harold. Random words, Toeplitz determinants and integrable systems. II. Advances in nonlinear mathematics and science. *Phys. D*, **152/153** (2001), 199–224.

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

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

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

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