Limiting Curves for IID Records

Jean-Dominique Deuschel *
ETH-Zentrum
CH-8092 Zurich
Switzerland

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

June 24, 1993; Revised October 12, 1993, May 1, 1994 and August 26, 1994

Abstract We consider the concentration of measure for n iid, two-dimensional random variables, under the conditioning that they form a record. Under mild conditions, we show that all random variables tend to concentrate, as $n \to \infty$, around limiting curves, which are the solutions of an appropriate variational problem. We also show that the same phenomenum occurs, without the records conditioning, for the longest increasing subsequence in the sample.

AMS 1991 Subject Classification: Primary 60G70. Secondary 60F10.

Abbreviated Title: Curves for IID Records.

Keywords: Records. Longest increasing subsequence. Large Deviations.

*This work was partially supported by the Swiss National Foundation grant Nr. 21-298333.90
†This work was partially supported by a US–ISRAEL BSF grant, by the Swiss National Foundation grant Nr. 21-298333.90 and by the fund of promotion of research at the Technion.

1 Introduction

Let $z_\alpha \triangleq (x_\alpha, y_\alpha)$, $\alpha = 1, 2, \cdots, n$ be n i.i.d., $\mathbb{R}^2$ valued random variables, each distributed according to the law $P(x, y)$ on $[0,\ 1]^2$. For $\ell \le n$, we say that a subsequence $\{z_{i_j} : i_1, i_2, ..., i_\ell \subseteq \{1, 2, ..., n\}\ \} \subseteq \{z_\alpha\}$ forms an increasing subsequence, or a *record sequence of length* ℓ, if

$$x_{i_j} < x_{i_{j+1}},\ y_{i_j} < y_{i_{j+1}},\ j = 1, ..., \ell - 1.$$

(Note that we do not require that $i_j < i_{j+1}$). Let $\ell_{\max}(n)$ denote the length of the longest increasing subsequence. We say that $\{(x_\alpha, y_\alpha)\}$ form a *record* sequence (in short: *form a record*) if $\ell_{\max}(n) = n$. This is equivalent to the existence of a permutation $\pi : \{1, \cdots, n\} \to \{1, \cdots, n\}$ such that

$$x_{\pi(\alpha)} < x_{\pi(\alpha+1)}, \quad y_{\pi(\alpha)} < y_{\pi(\alpha+1)}, \qquad \forall\, \alpha = 1, \cdots, n-1\,.$$

We are interested in the concentration of the measure P^n under the conditioning that a record exists. This question is motivated by the following result, due to Goldie and Resnick [5]. Let $Z_i \triangleq (X_i, Y_i)$, $i = 1, 2, \ldots$ be a sequence of i.i.d. $\mathbb{R}^2$ valued random variables, each distributed according to the law $\mathcal{P}(x, y)$ on $[0,\ 1]^2$. Let $L_o = 0$ and

$$L_n = \begin{cases} \infty, & \text{if } L_{n-1} = \infty \\ \inf\{m > L_{n-1} : X_m > \max_{j=1}^{m-1} X_j, Y_m > \max_{j=1}^{m-1} Y_j\}, & \text{if } L_{n-1} < \infty\,. \end{cases} \tag{1}$$

Let

$$R_n = \begin{cases} (1, 1), & \text{if } L_n = \infty, \\ (X_{L_n}, Y_{L_n}), & \text{if } L_n < \infty\,. \end{cases} \tag{2}$$

L_n is the sequence of simultaneous record times, while R_{L_n} denotes the record (extreme) values. Define $N = \sup\{n : L_n < \infty\}$, and, with $z = (x, y)$, define the hazard measure

$$H(dz) = \frac{\mathcal{P}(dz)}{1 - \mathcal{P}((0, x] \times (0, y])}\,.$$

If $H([0, 1]^2) < \infty$, then $N < \infty$ a.s. (see [4]). This is the case if $\mathcal{P}$ possesses an atom in $(1, 1)$ or, more generally, if $\mathcal{P}$ is obtained from a general distribution on $\mathbb{R}_+^2$ which charges $((0, 1)^2)^c$.

Suppose now that one conditions on N being large. One might ask what form does the record sequence take under this conditioning. Let $P(dz) = H(dz)/H([0, 1])^2$. A basic result in [5] (Proposition 2.2) states that if H possesses a bounded density on $[0, 1]^2$, and if the measure P^n concentrates, under the conditioning that a record exists, around a deterministic curve as n increases, then so do the values of R_i, $i = 1, 2, \ldots, N$, conditioned on $N = n$. Thus, the concentration of P^n under the record conditioning, which is the main object studied in this paper, plays a decisive role in understanding the structure of multi-dimensional extremal sequences. [5] contains also some results related to those in this paper.

It should be noted that the case of $P(x, y)$ uniform, or, more generally, independent x and y coordinates, is rather straight forward (see remarks 2,3 following Theorem 1). The general case however does require some work.

A naturally related question, suggested by Cochand, is the following: Let $\ell_{\max}(n)$ denote, as before, the length of the longest increasing subsequence. One may ask about the asymptotics of $\ell_{\max}(n)$ and the shape of the longest increasing subsequence. For the uniform case, this question is equivalent to the one tackled by Vershik and Kerov in [9], where they prove that $\ell_{\max}(n)/\sqrt{n} \to_{n\to\infty}$ 2 in probability, and it is not hard to see that in that case the longest increasing subsequence concentrates along the diagonal. It is of interest to extend their result to general densities. As we will show, the solution to this problem is intimately related to the solution of the records conditioning problem.

The organization of the paper is as follows: in the next section, we state our main results and prove that under mild conditions, records concentrate around limiting curves which form the solution of a variational problem. Section 3 studies the properties of the latter variational problem. In particular, we prove existence of absolutely continuous optimizing curves, and provide a characterization for those. Section 4 is devoted to examples. Finally, Section 5 deals with the longest increasing subsequence problem.

2 Main results

We will work here under the following hypotheses

(A1) $P(x,y)$ possesses a bounded density $p(x,y)$ with respect to Lebesgue measure in $[0,\,1]^2$.

(A2) $p(x,y)$ is C_b^1 and bounded below in $[0,\,1]^2$.

Before stating the last assumption we need some further notation: Let $B^{\nearrow}$ be the set of *non-decreasing*, right continuous functions $\phi : [0,1] \to [0,1]$. For $\phi \in B^{\nearrow}$, we have $\phi(t) = \int_0^t \dot{\phi}(s)ds + \phi_s(t)$ where ϕ_s is singular (and possesses a zero derivative almost everywhere). Next define $J : B^{\nearrow} \to \mathbb{R}^+$

$$J(\phi) = \int_0^1 \sqrt{\dot{\phi}(x)p(x,\phi(x))}\,dx \tag{3}$$

and denote by $K(J) \subseteq B^{\nearrow}$ the set of solutions to the variational problem

$$\overline{J} \triangleq \sup_{\phi\in B^{\nearrow}} J(\phi). \tag{4}$$

Our third assumption is

(A3) $K(J)$ is a finite set $\{\phi_1, ..., \phi_k\}$.

We claim the:

Theorem 1 *Under* **(A1)**–**(A3)**, *for each* $\delta > 0$,

$$\lim_{n\to\infty} P(\min_{\ell=1,\ldots,k}\ \max_{\alpha=1,\ldots,n} |y_\alpha - \phi_\ell(x_\alpha)| < \delta \ |\{(x_\alpha,y_\alpha)\}_{\alpha=1}^n \textit{ form a record }\ ) = 1\,. \tag{5}$$

Remarks: 1) We shall prove in the next section, that, under **(A1)** and **(A2)**, $K(J)$ is a non-empty compact subset of C_b^1. Actually we believe that **(A3)** follows from **(A1)** and **(A2)**, but don't know how to prove it. We show in Theorem 3, that each $\phi \in K(J)$ solves the boundary value problem

$$\ddot{\phi}(x) = \frac{p_x(x,\phi(x))}{p(x,\phi(x))}\dot{\phi}(x) - \frac{p_y(x,\phi(x))}{p(x,\phi(x))}\dot{\phi}(x)^2 \tag{6}$$

with boundary conditions

$$\phi(0) = 0, \quad \phi(1) = 1.$$

Also any two different solutions ϕ_1, ϕ_2 of (4) can only intersect at $x = 0$ and $x = 1$.
2) Theorem 1 admits a rather elementary proof in the case that P is the uniform law on $[0,1]^2$. Indeed, in that case, the coordinates are independent, and a record occurs if the two independent rearrangements of the $\{x_\alpha\}$ and $\{y_\alpha\}$ coincide (we note that this event has probability $1/n!$). Let $B^X_{k,\delta,n}$ denote the event that for all k, the $k-th$ smallest sample (out of n) of the x coordinate is in $[(k/n-\delta)\wedge 0, (k/n+\delta)\vee 1]$, let $B^Y_{k,\delta,n}$ denote the analogous event in the y coordinate, and let $B_{k,\delta,n} = B^X_{k,\delta,n} \cap B^Y_{k,\delta,n}$. If one can show that $P(\cap_k B_{k,n,\delta} | \ell_{\max}(n) = n) \to 1$ as $n \to \infty$, then one immediately deduces that the record must concentrate around the diagonal in the sense of Theorem 1. Note that conditioning on the occurence of a record amounts to conditioning on the ordering permutation being the same in both coordinates, an event which is independent of all $B_{k,\delta,n}$. Thus,

$$P(\cup_k B^c_{k,n,\delta} | \ell_{\max}(n) = n) = P(\cup_k B^c_{k,n,\delta}) \le P(\cup_k {B^X_{k,n,\delta}}^c) + P(\cup_k {B^Y_{k,n,\delta}}^c) = 2P(\cup_k {B^X_{k,n,\delta}}^c) \tag{7}$$

Note that

$$P(\cup_k {B^X_{k,n,\delta}}^c) \le P(\sup_{x\in[0,1]} |\frac{1}{n}\sum_{\alpha=1}^n 1_{[0,x]}(x_\alpha) - x| > \delta) \to_{n\to\infty} 0\,,$$

where the last limit follows from the Glivenko-Cantelli theorem (see, e.g., [7, Pg. 7]). Combining this with (7), one arrives at the desired conclusion. As will be seen in Section 4, the diagonal is also the unique solution of the optimization problem (4).

3) The case of $p(x,y) = p(x)q(y)$ can also be easily settled by a monotone change of coordinates which reduces the problem to the uniform case.

The following extension of the Vershik–Kerov theorem was conjectured by Cochand.

Theorem 2 *1. Assume* **(A1), (A2)**. *Then*

$$\lim_{n\to\infty} \ell_{\max}(n)/\sqrt{n} = 2\bar{J}\,, \tag{8}$$

where the limit is in probability.

2. Further assume **(A3)**. *Then, for each $\delta > 0$, and each record sequence $\{z_{i_\alpha}\}_{\alpha=1}^{\ell_{\max}(n)}$ of length $\ell_{\max}(n)$,*

$$\lim_{n\to\infty} P(\min_{j=1,\dots,k} \max_{\alpha=1,\dots,\ell_{\max}(n)} |y_{i_\alpha} - \phi_j(x_{i_\alpha})| < \delta\ ) = 1\,. \tag{9}$$

Remarks: 1) Actually, the result of Vershik and Kerov is stated in terms of the longest increasing subsequence of a random permutation. However, this is equivalent to our problem since, the x and y coordinates being independent in the case of uniform distribution, one may first re-arrange the x coordinate and ask for the longest increasing subsequence in the independent y coordinate, which is equivalent to the random permutation problem. The same argument applies to the general case of independent coordinates.
2) Theorem 2 extends naturally to the d-dimensional hypercube. Essentially the same argument shows that then, $\ell_{\max}/n^{1/d} \to c_d \overline{J}$, where c_d is the limiting constant for the uniform case. It is, to the best of our knowledge, an open and challenging problem to compute c_d.

As is often the case in limit theorems involving conditioning, the proof of Theorem 1 is split into lower and upper bounds. We use below the notation

$$\Omega \supset \Omega_n = \{\omega : \ \{(x_\alpha, y_\alpha)\}_{\alpha=1}^n \text{ form a record } \}$$

Lemma 1 *Assume* **(A1)–(A2)**. *Let $\phi : [0,1] \to [0,1]$ be a monotone, non decreasing, C_b^1 function. Then, for each $\delta > 0$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \left[e^{n\log\left(\frac{n}{e}\right)} P(\Omega_n, \max_{\alpha=1,\ldots,n} |y_\alpha - \phi(x_\alpha)| \le \delta) \right] \ge 2 \log J(\phi) \,. \tag{10}$$

Before proving the Lemma, let us recall a well known fact. Let K_x, K_y be integers and set $\Delta x = 1/K_x$, $\Delta y = 1/K_y$,

$$\Delta x_i = [(i-1)\Delta x, i\Delta x) \qquad \Delta y_j = [(j-1)\Delta y, j\Delta y),$$

$i = 1, ..., K_x, j = 1, ..., K_y$. Next denote by $\ell_n(i,j)$ the block empirical measure of $(z_1, ..., z_n)$:

$$\ell_n(i,j) = \frac{1}{n} \ |\{(x_\alpha, y_\alpha) : \ z_\alpha \in \Delta x_i \times \Delta y_j\}|$$

where $|\cdot|$ denotes the cardinality of a set.

In the sequel, we refer to any probability vector with weights which are integer multiples of $1/n$ as a *type*. For a given type μ on the above grid, consider a partition $\{M_{ij}(\mu)\}$ of $\{1, ..., n\}$ into disjoint connected components such that $|M_{ij}(\mu)| = n_{ij}(\mu) = n\mu_{i,j}$ and set $M_i(\mu) = \cup_j M_{ij}(\mu)$, $n_i(\mu) = |M_i(\mu)|$. Let us introduce a new family of random variables $\{\bar{z}_i = (\bar{x}_i, \bar{y}_i)\,, \ i = 1, ..., n\}$ defined by the law

$$P(d\bar{z}_1, ..., d\bar{z}_n) = \sum_{\mu \text{ is a type}} P(\ell_n = \mu) P_\mu(d\bar{z}_1, ..., d\bar{z}_n),$$

where

$$P_\mu(d\bar{z}_1, ..., d\bar{z}_n) = \prod_{i,j} \prod_{k \in M_{ij}(\mu)} \bar{P}_{ij}(d\bar{z}_k), \qquad \bar{P}_{ij}(d\bar{z}_k) = \bar{p}_{ij}(\bar{z}_k) d\bar{z}_k$$

with

$$\bar{p}_{ij}(z) = \frac{p(z)}{\int_{\Delta x_i} \int_{\Delta y_j} p(x,y)\, dx\, dy} 1_{z \in \Delta x_i \times \Delta y_j} \,.$$

Thus the random variables $\{\bar{z}_i\}$ are obtained from two subsequent drawings: first choose the block empirical measure ℓ_n according to the original law P and then distribute the random variables $\bar{z}$ inside the boxes $\Delta x_i \times \Delta y_j$ according to $\bar{P}_{ij}$.

Finally define the two empirical measures L_n and $\bar{L}_n$

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{z_i} \qquad \bar{L}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\bar{z}_i}.$$

An important step in the proof of Lemma 1 is the following intuitive, well known result which, for completeness, is proved in the Appendix:

Lemma 2 *For all measurable $A \subseteq \mathcal{M}_1([0,1]^2)$ (the set of probability distributions on $[0,1]^2$), we have*

$$P(L_n \in A) = P(\bar{L}_n \in A).$$

We can now proceed with the proof of Lemma 1:

Proof of Lemma 1:

Let δ_1 be such that $1/\delta_1$ is an integer, and let $\beta(\delta_1) = \beta > 4$ be a δ_1–dependent integer whose value will be fixed below. Choose $K_x = K_x(\delta_1)$ large enough such that $\Delta x < \delta$ and

$$\sup_{x_o \in [0,1)} \quad \sup_{y_o \in [0,1)} \frac{\displaystyle\sup_{\substack{|x-x_o|<\Delta x \\ |y-y_o|<\max(1,\sup_{x\in[0,1]}\dot{\phi}(x))\Delta x}} p(x,y)}{\displaystyle\inf_{\substack{|x-x_o|<\Delta x \\ |y-y_o|<\max(1,\sup_{x\in[0,1]}\dot{\phi}(x))\Delta x}} p(x,y)} \leq (1+\delta_1) \tag{11}$$

(This is possible due to (**A2**)). Let now

$$\Delta \stackrel{\triangle}{=} \max_{i=1}^{K_x} \sup_{s,t\in\Delta x_i} (|\dot{\phi}(s) - \dot{\phi}(t)| \vee |\phi(s) - \phi(t)|)$$

(Δ is finite and $\Delta \underset{\Delta x \to 0}{\longrightarrow} 0$ since ϕ is C_b^1). Increase K_x if necessary to make

$$\Delta < \min(\delta/\beta, \delta_1/2)\,. \tag{12}$$

Choose now $K_y = \beta K_x$. Recall that by the fact that the $\{(x_\alpha, y_\alpha)\}_{\alpha=1}^n$ forms an i.i.d. sequence, for any *type* μ,

$$\frac{e^{-nH(\mu|\hat{P})}}{(n+1)^{K_x K_y}} \leq P(\ell_n = \mu) \leq e^{-nH(\mu|\hat{P})} \tag{13}$$

where $H(\mu|\hat{P}) = \sum_{i,j} \mu_{ij} \log \frac{\mu_{ij}}{\hat{P}_{ij}}$ is the relative entropy of μ with respect to $\hat{P}$, and $\hat{P}_{ij} = \int_{\Delta y_i}\int_{\Delta x_i} p(x,y)\,dx\,dy$. (See [2], Lemma 2.1.9). For a given measure μ on $\{1,\ldots,K_x\} \times \{1,\ldots,K_y\}$, let

$$B_i(\mu) = \min\{j : \mu_{ij} > 0\}, \quad T_i(\mu) = \max\{j : \mu_{ij} > 0\}$$

with $T_0(\mu) = 0$ and $B_i(\mu) = T_i(\mu) = T_{i-1}(\mu)$ if $\mu_{ij} = 0 \; \forall j$. It is obvious that when $\ell_n = \mu$ then the support of $\{L_n\}$ belongs to the set $\cup_i(\Delta x_i \times [(B_i(\mu) - 1)\Delta y, T_i(\mu)\Delta y))$.

Define the events

$$
\begin{aligned}
r_n = r_n(L_n) &\stackrel{\triangle}{=} \{\omega : \; \forall i = 1, \cdots, K_x - 1, B_{i+1}(\ell_n) > T_i(\ell_n)\} \cap \Omega_n \\
R_n = R_n(L_n) &\stackrel{\triangle}{=} \{\omega : \; \forall i = 1, \cdots, K_x - 1, B_{i+1}(\ell_n) \geq T_i(\ell_n)\} \cap_i \Omega_n^i \, ,
\end{aligned}
$$

where $C^i = \{\alpha : x_\alpha \in [(i-1)\Delta x, i\Delta x)\}$ and

$$
\Omega_n^i = \{\omega : \{(x_\alpha, y_\alpha)\}_{\alpha \in C^i} \text{ form a record}\} \, .
$$

Clearly, $r_n \subseteq \Omega_n \subseteq R_n$.

Finally, denote

$$
E^\delta = E^\delta(L_n) = \{\omega : \; |y_\alpha - \phi(x_\alpha)| \leq \delta, \quad \forall \alpha = 1, \cdots, n\} \, ,
$$

and write $\bar{E}^\delta = E^\delta(\bar{L}_n)$, $\bar{r}_n = r_n(\bar{L}_n)$. Now, by Lemma 2

$$
\begin{aligned}
P(E^\delta \cap \Omega_n) &\geq P(E^\delta \cap r_n) = P(\bar{E}^\delta \cap \bar{r}_n) = \sum_{\mu \text{ type}} P(\ell_n = \mu) P_\mu(\bar{E}^\delta \cap \bar{r}_n) \\
&\geq \sum_{\mu \in A(\delta)} P(\ell_n = \mu) P_\mu(\bar{r}_n) \qquad (14)
\end{aligned}
$$

where

$$
A(\delta) = \{\mu \text{ type } : B_{i+1}(\mu) > T_i(\mu), \text{ and } \; \mu_{i,j} = 0 \text{ if } |\phi(i\Delta x) - j\Delta y| > \frac{\delta}{2}, \, j = 1, ..., K_y, \, i = 1, ..., K_x\}
$$

(we used in the last inequality the fact that $\Delta y + \Delta < \delta/2$). Let

$$
\bar{r}_n^i = \{\bar{z} = \{\bar{z}_\alpha\}_{\alpha \in M_i} : \text{ form a record } \} \, ,
$$

then for a type $\mu \in A(\delta)$, by construction we have from (11)

$$
P_\mu(\bar{r}_n) = P_\mu(\bigcap_i \bar{r}_n^i) = \prod_i P_\mu(\bar{r}_n^i) \geq \Big(\frac{1}{1+\delta_1}\Big)^n \prod_i P_\mu^u(\bar{r}_n^i)
$$

where P_μ^u denotes the distribution corresponding to $\bar{P}_{ij}$ being the uniform distribution on each $\Delta x_i \times \Delta y_j$.

Lemma 3 *For each i,*

$$
P_\mu^u(\bar{r}_n^i) = \frac{1}{n_i(\mu)!} \, . \qquad (15)
$$

Proof: Since μ is fixed, let us drop the dependence of μ in the formulae. Without loss of generality, we may assume that $M_i = \{1, 2, ..., n_i\}$ and $M_{ij} = \{m_{j-1}+1, ..., m_j\}$ where $m_0 = 0$, $m_j = \sum_{k=1}^{j} n_{ik}$, $j = 1, ..., J^i = T_i - B_i + 1$.

Let Π_s^i = { permutations $\pi : M_i \to M_i : \pi(M_{ij}) = M_{ij}, j = 1, ..., J^i\}$. Now

$$\begin{aligned}
P_\mu^u(r_n^i) &= \textstyle\sum_{\pi\in\Pi_s^i} P_\mu^u(\bar{x}_{\pi(1)} < ... < \bar{x}_{\pi(n_i)}, \bar{y}_{\pi(1)} < ... < \bar{y}_{\pi(n_i)}) \\
&= \textstyle\sum_{\pi\in\Pi_s^i} P_\mu^u(\bar{x}_{\pi(1)} < ... < \bar{x}_{\pi(n_i)}, \bar{y}_{\pi(1)} < ... < \bar{y}_{\pi(m_1)}, \bar{y}_{\pi(m_1+1)} < ... < \bar{y}_{\pi(m_2)}, \\
&\qquad ..., \bar{y}_{\pi(m_{J^i-1}+1)} < ... < \bar{y}_{\pi(m_{J^i})}) \\
&= |\Pi_s^i| P_\mu^u(\bar{x}_1 < ... < \bar{x}_{n_i}, \bar{y}_1 < ... < \bar{y}_{m_1}, \bar{y}_{m_1+1} < ... < \bar{y}_{m_2}, ..., \bar{y}_{m_{J^i-1}+1} < ... < \bar{y}_{m_{J^i}}) \\
&= |\Pi_s^i| P_\mu^u(\bar{x}_1 < ... < \bar{x}_{n_i}) \textstyle\prod_{j=1}^{J^i} P_\mu^u(\bar{y}_{m_{j-1}+1} < ... < \bar{y}_{m_j}) \\
&= |\Pi_s^i| \tfrac{1}{n_i!} \textstyle\prod_{j=1}^{J^i} \tfrac{1}{n_{ij}!} = \tfrac{1}{n_i!}.
\end{aligned}$$

$\square$

Combining (13), (14) and (15), one arrives at

$$P(E^\delta, \Omega_n) \geq \left(\frac{1}{1+\delta_1}\right)^n \prod_{i=1}^{K_x} \frac{1}{n_i(\mu)!} e^{-nH(\mu|\hat{P})} \frac{1}{(n+1)^{K_x K_y}}$$

for any $\mu \in A(\delta)$. Write $\mu_i = n_i(\mu)/n$. Now, by Stirling's formula ($n! \leq \left(\frac{n+1}{e}\right)^{n+1} \cdot e$) one obtains, for n large enough,

$$\begin{aligned}
P(E^\delta, \Omega_n) \quad &\geq \quad \frac{1}{2}\left(\frac{1}{1+\delta_1}\right)^n \frac{1}{(n+1)^{K_x K_y}} e^{-K_x K_y} \cdot \left(\frac{1}{n+1}\right)^{K_x} \\
&\qquad \cdot e^{-n\log\left(\frac{n}{e}\right)} \cdot e^{-n\sum_{ij}\mu_{ij}\log\frac{\mu_{ij}}{\hat{P}_{ij}}} \\
&\qquad \cdot e^{-n\sum_i \mu_i \log \mu_i} \\
&\triangleq \quad g_n e^{-n\log\left(\frac{n}{e}\right)} e^{-n\sum_{ij}\mu_{ij}\log\frac{\mu_{ij}}{\hat{P}_{ij}}} \cdot e^{-n\sum_i \mu_i\log\mu_i} \qquad (16)
\end{aligned}$$

where μ satisfies the support constraint, and, for each $\beta = \beta(\delta_1)$ independent of n, $\lim_{\delta_1\to 0}\lim_{n\to\infty}\frac{1}{n}\log g_n = 0$.

Our strategy in the sequel is to find appropriate μ which, together with (16), will yield the required lower bound. To this end, it is convenient to first fix the support limits B_i, T_i and then look for a μ such that $B_i(\mu) = B_i$ and $T_i(\mu) = T_i$. Let

$$\mu_{ij} = \begin{cases} \frac{\mu_i \hat{P}_{ij}}{\sum_{j=B_i}^{T_i}\hat{P}_{ij}} & j \in \{B_i, \cdots, T_i\} \\ 0 & \text{elsewhere} \end{cases}$$

Then,

$$\sum_{ij}\mu_{ij}\log\frac{\mu_{ij}}{\hat{P}_{ij}} + \sum_i \mu_i\log\mu_i = 2\sum_i \mu_i \log\frac{\mu_i}{\sqrt{\sum_{j=B_i}^{T_i}\hat{P}_{ij}}}. \qquad (17)$$

Choosing $\mu_i = c\sqrt{\sum_{j=B_i}^{T_i} \hat{P}_{ij}}$ where c is such that $\sum_i \mu_i = 1$ yields

$$2\sum_i \mu_i \log \frac{\mu_i}{\sqrt{\sum_{j=B_i}^{T_i} \hat{P}_{ij}}} = 2\log c = -2\log\Big(\sum_i \sqrt{\sum_{j=B_i}^{T_i} \hat{P}_{ij}}\Big). \tag{18}$$

(Note that this choice of μ_{ij} and μ_i is optimal in that it maximizes the right hand side of the expression (16)).

To conclude the proof of Lemma 1, assume first that $\gamma = \inf_{x\in[0,1]} \dot{\phi}(x) > 0$, and let

$$B_i = \left\lceil \frac{\phi(i\Delta x)}{\Delta y} \right\rceil, \quad T_i = B_{i+1} - 1\,.$$

Note that for $B_i \le j \le T_i$, one has by (11) that

$$\frac{1}{1+\delta_1} \le \frac{\hat{P}_{ij}}{p(i\Delta x, \phi(i\Delta x)) \cdot \Delta x \Delta y}\,.$$

Let $\beta = \beta(\delta_1)$ be large enough such that simultaneously, $\phi(i\Delta x) \in [\Delta y B_i, \Delta y T_i)$, and

$$\frac{1}{1+\delta_1} \le \frac{T_i - B_i + 1}{\dot{\phi}(i\Delta x) \cdot \frac{\Delta x}{\Delta y}}\,.$$

(This is possible since $\gamma > 0$ by using (12)!). Hence,

$$\begin{aligned} \lim_{\delta_1\to 0} 2\log\Big(\textstyle\sum_i \sqrt{\sum_{j=B_i}^{T_i} \hat{P}_{ij}}\Big) &\ge \lim_{\Delta x\to 0} 2\log\Big(\textstyle\sum_i \sqrt{p(i\Delta x, \phi(i\Delta x))\dot{\phi}(i\Delta x)}\,\Delta x\Big) \\ &= 2\log\Big(\textstyle\int_0^1 \sqrt{p(x,\phi(x))\dot{\phi}(x)}\,dx\Big)\,. \end{aligned} \tag{19}$$

If μ were a type for each δ_1, (19) together with (16), (17) and (18) would yield Lemma 1. The general case follows by approximation using the continuity of the RHS of (16) in μ_{ij} by noting that, for every large enough n, one may find a type μ_n arbitrarily close to μ with same support limits B_i, T_i. Finally, the case $\gamma = 0$ follows by considering a sequence of ϕ_m such that $\dot{\phi}_m(x) > 0$ for all $x \in [0,1]$ while $\dot{\phi}_m$ converges uniformly to $\dot{\phi}$ on $[0,1]$. □

We prove in the next section, cf. Corollary 1, that **(A1)**–**(A3)** imply

(A4) Let $||\cdot||$ denote the supremum norm on $[0,1]$. For every $\delta > 0$ there exists an $\epsilon(\delta) > 0$ such that any piecewise linear, non-decreasing $\phi : [0,1] \to [0,1]$ with $||\phi - \phi_\ell|| > \delta$ for $\ell = 1, 2, \ldots, k$ satisfies $J(\phi) \le \overline{J} - \epsilon(\delta)$.

Lemma 4 *Assume* **(A1)**–**(A4)**. *Let $\delta > 0$, small enough, be given. Then,*

$$\limsup_{n\to\infty} \frac{1}{n} \log e^{n\log(\frac{n}{e})} P(\Omega_n, \min_{\ell=1,\ldots,k} \max_{\alpha=1,\ldots,n} |y_\alpha - \phi_\ell(x_\alpha)| > \delta\,) \le 2\log(\bar{J} - 2\epsilon(\delta/4))$$

where $\epsilon(\delta) > 0$.

Proof: It follows from Theorem 4 below that we may, and will, consider only $\phi_\ell \in C_b^1$, and furthermore that $\dot{\phi}_\ell(x) > 0$ for all $x \in (0,1)$. Fix $\delta_1 > 0$, and let $\delta_2 = \delta_2(\delta_1) \le 1$ be such that

$$\sup_{x_o \in [0,1)} \quad \sup_{y_o \in [0,1)} \quad \frac{\sup\limits_{\substack{|x-x_o|<\delta_2 \\ |y-y_o|<\delta_2}} p(x,y)}{\inf\limits_{\substack{|x-x_o|<\delta_2 \\ |y-y_o|<\delta_2}} p(x,y)} \le (1+\delta_1)\,, \tag{20}$$

let $K_x = [\delta_2^{-3}]$ and $\Delta x = K_x^{-1}$. Reduce δ_2 if necessary to have also $\Delta x < \delta/(2\max_{x\in[0,1]} \dot{\phi}_\ell(x))$, $\ell = 1, \dots, k$. Fix $\beta > 4$ (eventually, a limit in $\beta \to \infty$ will be taken). Define next $K_y = \beta K_x$ and $\Delta y = K_y^{-1}$ as in the proof of Lemma 1.

Let

$$\tilde{E}^\delta = \tilde{E}^\delta(L_n) = \{\omega : \min_{\ell=1,\dots,k} \max_{\alpha=1,\dots,n} |y_\alpha - \phi_\ell(x_\alpha)| > \delta\}$$

and for fixed sequences $\underline{t} = (t_1, \cdots, t_{K_x}), \underline{b} = (b_1, \cdots, b_{K_x})$ which satisfy $b_{i+1} \ge t_i$, set

$$F^\delta(\underline{t}, \underline{b}) = \{\omega : \ T_i(\ell_n) = t_i, B_i(\ell_n) = b_i, \quad \min_{\ell=1,\dots,k} \max_i \max_{j_\ell \in [b_i, t_i]} |\phi_\ell(i\Delta x) - j_\ell \Delta y| > \delta/2, \ \}$$

Then using R_n instead of r_n, (16) is replaced, for every sequence $(\underline{t}, \underline{b})$, by

$$\begin{aligned} P(R_n, F^\delta(\underline{t}, \underline{b})) \ &\le \sum_{\mu \text{ type}: T_i(\mu)=t_i, B_i(\mu)=b_i} P(\ell_n = \mu) P_\mu(\cap_i \overline{r}_n^i) \\ &\le \ (1+\delta_1)^n e^{-n\log\frac{n}{e}} \sum_{\mu \text{ type}: T_i(\mu)=t_i, B_i(\mu)=b_i} e^{-n\sum_{ij} \mu_{ij} \log \frac{\mu_{ij} \sum_{k=b_i}^{t_i} \mu_{ik}}{\hat{P}_{ij}}} . \end{aligned}$$

(Note that $n! \ge (n/e)^n$, and use the upper bound in (13).) For each (K_x, K_y) let

$$\mathcal{F} = \{(\underline{t}, \underline{b}) : \text{ for all } \ell = 1, \dots, k, \ \exists j_\ell \in [b_i, t_i] \ , |\phi_\ell(i\Delta x) - j_\ell \Delta y| > \delta/2, b_{i+1} \ge t_i \ \ i = 1, ..., K_x\}\,.$$

Note that $|\mathcal{F}| \le K_y^{2K_x}$. Note also that since $\Delta x < \delta/(2\max_{x\in[0,1]} \dot{\phi}_\ell(x))$, one has that $R_n \cap \tilde{E}^\delta \subset \bigcup_{(\underline{t},\underline{b})\in\mathcal{F}} R_n \cap F^\delta(\underline{t}, \underline{b})$. Thus,

$$P(\Omega_n, \tilde{E}^\delta) \le P(R_n, \tilde{E}^\delta) \le \overline{g}_n e^{-n\log\frac{n}{e}} \sup_{(\underline{t},\underline{b})\in\mathcal{F}} \ \sup_{\{\mu : \text{support } \mu \in \cup_i\{i,[b_i,t_i]\}} e^{-n\sum_{ij} \mu_{ij} \log \frac{\mu_{ij} \sum_{k=b_i}^{t_i} \mu_{ik}}{\hat{P}_{ij}}}$$

where $\lim_{\delta_1\to 0} \lim_{n\to\infty} \frac{1}{n} \log \overline{g}_n = 0$, uniformly in all other parameters.

Fixing (b_i, t_i) and using Lagrange multipliers to optimize over μ which satisfies the constraints, one obtains

$$P(R_n, \tilde{E}^\delta) \le \overline{g}_n e^{-n\log\frac{n}{e}} \sup_{(\underline{t},\underline{b})\in\mathcal{F}} e^{2n\log\sum_i \sqrt{\hat{P}_i^{(b_i,t_i)}}} \tag{21}$$

where $\hat{P}_i^{(b_i,t_i)} = \sum_{j=b_i}^{t_i} \hat{P}_{ij}$. For each $(\underline{t}, \underline{b}) \in \mathcal{F}$, let

$$\Theta_1(\underline{b},\underline{t}) = \{i : K_y^{-1}(b_{i+1} - b_i + 1) < \delta_2\}, \;\; \Theta_2(\underline{b},\underline{t}) = \{i : 1 \geq K_y^{-1}(b_{i+1} - b_i + 1) \geq \delta_2\}\,.$$

Note that

$$\sum_{i \in \Theta_2(\underline{b},\underline{t})} \sqrt{\hat{P}_i^{(b_i,t_i)}} \leq \frac{\sqrt{\Delta x \sup_{x,y\in[0,1]} p(x,y)}}{\delta_2} \leq \sqrt{\delta_2 \sup_{x,y\in[0,1]} p(x,y)}\,,$$

where we have in the first inequality used the fact that $|\Theta_2(\underline{b},\underline{t})| \leq 1/\delta_2$. On the other hand, for $i \in \Theta_1(\underline{b},\underline{t})$,

$$\frac{1}{\delta_1 + 1} \leq \frac{\hat{P}_{ij}}{\Delta x \Delta y\, p(i\Delta x, b_i \Delta y)} \leq 1 + \delta_1$$

for any $j \in \{b_i, b_i + 1, \ldots, t_i\}$. Thus,

$$\begin{aligned}
\sum_i \sqrt{\hat{P}_i^{(b_i,t_i)}} &\leq \sqrt{(1+\delta_1)} \sum_{i\in\Theta_1(\underline{b},\underline{t})} \sqrt{p(i\Delta x, b_i\Delta y)(t_i - b_i + 1)\Delta y \Delta x} \\
&\quad + \sqrt{\delta_2 \sup_{x,y\in[0,1]} p(x,y)} \\
&\leq (1+\delta_1) \int_0^1 \sqrt{p(x,\phi_p(x))(\dot{\phi}_p(x) + \beta^{-1})}dx \\
&\quad + \sqrt{\delta_2 \sup_{x,y\in[0,1]} p(x,y)}\,,
\end{aligned}$$

where $\phi_p(\cdot)$ denotes the polygonal, monotone curve formed by $(i\Delta x, b_i\Delta y)$. Note that, for each $\ell = 1, \ldots, k$, $||\phi_p - \phi_\ell|| > \delta/4$. It follows that

$$\begin{aligned}
\sum_i \sqrt{\hat{P}_i^{(b_i,t_i)}} &\leq (1+\delta_1) \sup_{\{\phi\nearrow,\ \phi \text{ piecewise linear },\min_{\ell=1,\ldots,k} ||\phi-\phi_\ell||>\delta/4\}} \int_0^1 \sqrt{p(x,\phi(x))\dot{\phi}(x)}dx \\
&\quad +(1+\delta_1)\sqrt{\sup_{x,y\in[0,1]} p(x,y))}(\sqrt{\beta^{-1}} + \sqrt{\delta_2}) \\
&\leq (1+\delta_1)(\overline{J} - \epsilon(\delta/4)) + (1+\delta_1)\sqrt{\sup_{x,y\in[0,1]} p(x,y))}(\sqrt{\beta^{-1}} + \sqrt{\delta_2})
\end{aligned}$$

where the last inequality is due to **(A4)**. Using (21), taking first $n \to \infty$, then $\beta \to \infty$, followed by $\delta_2 \searrow 0$ and finally $\delta_1 \searrow 0$, yields Lemma 4. □

Proof of Theorem 1: Theorem 1 follows from Lemmas 1 and 4 by noting (see Corollary 1 below) that **(A4)** is implied by **(A1)-(A3)**. □

Remarks: 1. Let $\{S_{ij} = [a_i, a_{i+1}) \times [b_j, b_{j+1}), 1 \leq i,j \leq m\}$ be a finite partition of the unit square $[0,1)^2$ into rectangles, then an inspection of the proof reveals that one could replace **(A2)** by the following assumption:

(A2') The density $p(x,y)$ is C_b^1 and bounded below on each S_{ij}.

In this case, each $\phi \in K(J)$ is continuous in $[0,1)^2$ and piecewise C_b^1 in each S_{ij}. Note that now, different solutions may intersect inside $[0,1]^2$.

2. We could also allow that $p(x,y)$ vanishes on some S_{ij}:

(A2") On each square S_{ij}, the density $p(x,y)$ is either C_b^1 and bounded below, or $p(x,y)=0$.

Of course in this case we have to restrict the variational problem (2) to S^+, the set of squares S_{ij} where $p(x,y)>0$. The main difference with **(A2')**, is the discontinuity of the solutions. Under both **(A2')** and **(A2")**, two different solutions $\phi_1, \phi_2 \in K(J)$ may share the same line segment on some $S_{ij} \in S^+$, cf. the checkerboard example of Section 4.

3. Note that by rewriting ϕ as a parameterized curve, i.e. rewriting the curve $x \to (x, \phi(x))$ as $t \to \psi(t) = (\psi_1(t), \psi_2(t))$, J may be written more symmetrically as

$$J(\psi) = \int_0^1 \Big(p(\psi(t))\dot{\psi}_1(t)\dot{\psi}_2(t)\Big)^{1/2} dt.$$

Based on this expression and the proof of Theorem 1, one can reasonably expect that the results of this section extend to d-dimensional i.i.d. records with increasing curves $t \to \psi(t) = (\psi_1(t), \dots, \psi_d(t)) \in [0,1]^d$ maximizing

$$J(\psi) = \int_0^1 \Big(p(\psi(t))\dot{\psi}_1(t) \cdot \ldots \cdot \dot{\psi}_d(t)\Big)^{1/d} dt.$$

4. Based on the characterization of μ_{ij} as an empirical measure, one could actually prove a slightly stronger statement. Namely, let π be the permutation which reorders $\{x_\alpha\}$, and define the (random) piecewise constant function $g : [0,1] \to [0,1]$ by $g(t) = y_{\pi([nt])}$. Then, under **(A1)–(A3)**,

$$\lim_{n\to\infty} P\big(\min_{\ell=1,\dots,k} \sup_{t\in[0,1]} |g(t) - \phi_\ell(t)| < \delta | \{(x_\alpha, y_\alpha)\}_{\alpha=1}^n \text{ form a record } \big) = 1\,. \tag{22}$$

3 The variational problem

Let us first derive an existence result for the optimization problem under assumptions **(A1)-(A2)**.

Theorem 3 *Assume* **(A1)–(A2)**. *The optimization problem*

$$\overline{J} \triangleq \sup_{\phi \in B^{\nearrow}} \left(\int_0^1 \sqrt{\dot{\phi}(x) p(x, \phi(x))}\, dx \right) \tag{23}$$

possesses a solution over the space $B^{\nearrow}$.

Proof: Let first $\{\phi_n\} \subseteq B^{\nearrow}$ be a minimizing sequence of $-J$. Identify each $\phi \in B^{\nearrow}$ with a (positive) measure μ_ϕ on $[0,1]$, equipped with the Borel field, by $\mu_\phi([0,t]) = \phi(t)$. Equip the space of signed measure on $[0,1]$, denoted M, with the weak topology generated by $C[0,1]$, and note that both $M = C[0,1]^*$ and $C[0,1] = M^*$, where * denotes the topological dual. Next, note that by Helly's theorem, on a subsequence which is denoted again by $\{n\}$, $\phi_n \to \phi_\infty \in B^{\nearrow}$ a.e.. Thus,

$$|\int_0^1 \sqrt{p(x,\phi_n(x))\dot{\phi}_n(x)}dx - \int_0^1 \sqrt{p(x,\phi_\infty(x))\dot{\phi}_n(x)}dx| \leq c\sqrt{\int_0^1 |\phi_n(x) - \phi_\infty(x)|dx} \to_{n\to\infty} 0\,,$$

where c is a positive constant independent of ϕ_n. It therefore only remains to show that

$$\liminf_{n\to\infty} - \int_0^1 \sqrt{p(x,\phi_\infty(x))\dot{\phi}_n(x)}dx \geq -J(\phi_\infty)\,.$$

In order to proceed, we represent the functional above as an appropriate Legendre transform. This will yield the required lower semicontinuity. To this end, let $\mathcal{X} = C[0,1]$, $\mathcal{X}^* = M$ and $g(x) = p(x,\phi_\infty(x))$. As noted above, $\mathcal{X}^{**} = \mathcal{X}$. For any $\phi \in \mathcal{X}^*$, let

$$G(\phi) = \begin{cases} -\int_0^1 \sqrt{g(x)\dot{\phi}_{\mathrm{ac}}(x)}\,dx & \text{if } \phi \text{ is a nonnegative measure} \\ \infty & \text{otherwise} \end{cases}$$

where ϕ_{ac} denotes the absolutely continuous part of ϕ. For any $\psi \in \mathcal{X}$, define

$$\Lambda(\psi) = \begin{cases} \infty & \text{if } 1/\psi \notin L^1([0,1]) \text{ or } \psi \geq 0 \text{ on a set of positive Lebesgue measure} \\ -\frac{1}{4}\int_0^1 \frac{g(x)}{\psi(x)}dx & \text{otherwise} \end{cases}$$

For $\phi \in \mathcal{X}^*$, let

$$\Lambda^*(\phi) = \sup_{\psi\in\mathcal{X}} {}_{\mathcal{X}}\langle\psi,\phi\rangle_{\mathcal{X}^*} - \Lambda(\psi)\,.$$

where ${}_{\mathcal{X}}\langle\psi,\phi\rangle_{\mathcal{X}^*} = \int_0^1 \psi(x)\phi(dx)$ denotes the duality pairing between $\mathcal{X}$ and $\mathcal{X}^*$. Clearly, $\Lambda^*(\phi)$ is lower-semicontinuous. The existence theorem thus follows from the

Lemma 5

$$\Lambda^*(\phi) = G(\phi)\,.$$

Proof: Assume first that there exists a set A with $\phi(A) < 0$. Since ϕ is regular, one may find a sequence of continuous functions $0 \leq \psi_n \leq 1$ such that

$$\lim_{n\to\infty} {}_{\mathcal{X}}\langle\psi_n,\phi\rangle_{\mathcal{X}^*} = \phi(A) < 0\,.$$

Let $\Psi_n = -n\psi_n - 1$. It follows that

$$\lim_{n\to\infty} {}_{\mathcal{X}}\langle\Psi_n,\phi\rangle_{\mathcal{X}^*} - \Lambda(\Psi_n) = \infty\,.$$

We may thus assume that ϕ is non negative. Note that in this case, $\Lambda^*(\phi_{\mathrm{ac}}) \geq \Lambda^*(\phi)$. Let $\psi_n \in \mathcal{X}$ be a sequence such that

$$\lim_{n\to\infty} {}_{\mathcal{X}}\langle\psi_n,\phi_{\mathrm{ac}}\rangle_{\mathcal{X}^*} - \Lambda(\psi_n) = \Lambda^*(\phi_{\mathrm{ac}})\,,$$

with $c_n^{-1} \geq \psi_n \geq c_n > 0$. Let B be a Borel set such that $\phi_{\rm ac}([0,1] \setminus B) = \phi_{\rm ac}([0,1])$ and $\phi_{\rm s}([0,1]) = \phi_{\rm s}(B)$, where $\phi_{\rm s} = \phi - \phi_{\rm ac} \geq 0$. For each ϵ, let ψ_n^ϵ denote the ϵ continuous modification of $\psi_n 1_{B^c}$, that is $\psi_n^\epsilon = \psi_n 1_{B^c}$ on a set C with $\phi(C^c) < \epsilon, \mu(C^c) < \epsilon$ (μ denotes Lebesgue measure, and such a modification may be found by Lusin's theorem). Then, for some constant c,

$$\begin{aligned}
{}_{\mathcal{X}}\langle \psi_n, \phi_{\rm ac}\rangle_{\mathcal{X}^*} - \Lambda(\psi_n) &= {}_{\mathcal{X}}\langle \psi_n 1_{B^c}, \phi\rangle_{\mathcal{X}^*} - \Lambda(\psi_n 1_{B^c}) \\
&= {}_{\mathcal{X}}\langle \psi_n^\epsilon, \phi\rangle_{\mathcal{X}^*} + \frac{1}{2}\int_0^1 \frac{g}{\psi_n^\epsilon} dx + {}_{\mathcal{X}}\langle \psi_n - \psi_n^\epsilon, \phi\rangle_{\mathcal{X}^*} + \frac{1}{2}\int_0^1 g(\frac{1}{\psi_n} - \frac{1}{\psi_n^\epsilon}) dx \\
&\leq \Lambda^*(\phi) + \frac{\phi(C^c)}{c_n} + \frac{c}{c_n}\mu(C^c) \to_{\epsilon\to 0} \Lambda^*(\phi)
\end{aligned}$$

It follows that $\Lambda^*(\phi) = \Lambda^*(\phi_{\rm ac})$.

It remains to compute $\Lambda^*(\phi)$ for absolutely continuous, non decreasing ϕ. In the sequel, we implicitly assume that ϕ is absolutely continuous, and write

$$ {}_{\mathcal{X}}\langle \psi, \phi\rangle_{\mathcal{X}^*} = \int_0^1 \psi \dot\phi dx \, . $$

Note that, by a direct computation and [3], page 94, Theorem 2,

$$\begin{aligned}
G(\phi) &= \sup\{\int_0^1 \psi\dot\phi dx + \frac{1}{2}\int_0^1 \frac{g}{\psi} \, dx \, : \, \psi \text{ bounded, measurable, non positive}\} \\
&= \sup\{\int_0^1 \psi\dot\phi dx + \frac{1}{2}\int_0^1 \frac{g}{\psi} \, dx \, : \, \psi \text{ measurable, non positive}\} \, .
\end{aligned}$$

To complete the proof it therefore remains to check that

$$\begin{aligned}
&\sup\{\int_0^1 \psi\dot\phi dx + \frac{1}{2}\int_0^1 \frac{g}{\psi} \, dx \, : \, \psi \text{ bounded, measurable, non positive}\} \\
&= \sup\{\int_0^1 \psi\dot\phi dx + \frac{1}{2}\int_0^1 \frac{g}{\psi} \, dx \, : \, \psi \text{ continuous, non positive}\} \, .
\end{aligned}$$

This is again a straight forward application of Lusin's theorem. This completes the proof of the lemma and of the Theorem. ☐

Remark: As pointed out by the referee, one may, by the change of coordinates $y = \int_0^x \sqrt{g(t)}dt$ reduce the problem to the issue of lower semi-continuity of $G(\phi)$ for $g(\cdot) = 1$. Yuval Peres has shown us a direct proof of the latter fact, based on a construction of a suitable subsequence whose derivative converges almost everywhere.

While proving the existence of minimizers is somewhat involved, the characterization of minimizers is actually a consequence of the classical calculus of variations. Indeed, we have:

Theorem 4 *Assume* **(A1)–(A2)**. *Then, any optimizer* $\phi \in K(J)$ *of (4) is of class* C_b^1, *with absolutely continuous derivative, and satisfies the equation*

$$ \ddot\phi(x) = \frac{p_x(x,\phi(x))}{p(x,\phi(x))}\dot\phi(x) - \frac{p_y(x,\phi(x))}{p(x,\phi(x))}\dot\phi(x)^2 \tag{24} $$

with boundary conditions

$$\phi(0) = 0, \quad \phi(1) = 1$$

and $\dot{\phi} > 0$. Moreover different solutions to the variational problem can only intersect at $x = 0$ and $x = 1$.

Proof: We first show that it is enough to consider absolutely continuous minimizers with fixed boundary conditions.

Lemma 6 *Assume* **(A1)–(A2)***. Let $\phi \in K(J)$. Then, $\phi_s = 0$, $\phi(0) = 0$ and $\phi(1) = 1$, and ϕ cannot have an interval where it is constant.*

Proof: Assume first that $\phi_s \neq 0$. Then, in particular, there exists a sequence of intervals $[a_n, b_n]$ such that

$$\frac{\phi(b_n) - \phi(a_n)}{\phi_{\rm ac}(b_n) - \phi_{\rm ac}(a_n)} \to_{n\to\infty} \infty.$$

Let $\phi_n(t) = \phi(t)$ if $t \notin [a_n, b_n]$ and $\phi_n(t) = \phi(a_n) + (t - a_n)(\phi(b_n) - \phi(a_n))/(b_n - a_n)$ otherwise. Denoting $p = \min_{x,y} \sqrt{p(x,y)}, P = \max_{x,y} \sqrt{p(x,y)}$, it is easy to check that

$$\begin{aligned} J(\phi) - J(\phi_n) &\leq \int_{a_n}^{b_n} P\sqrt{\dot{\phi}(s)}ds - p\sqrt{(b_n - a_n)(\phi(b_n) - \phi(a_n))} \\ &\leq P\sqrt{(b_n - a_n)(\phi_{\rm ac}(b_n) - \phi_{\rm ac}(a_n)} \\ &\qquad -p\sqrt{(b_n - a_n)(\phi(b_n) - \phi(a_n))} < 0 \end{aligned}$$

for sufficiently large n, where the second inequality is a consequence of Jensen's inequality and the first of the definition of a_n, b_n. It follows that ϕ cannot possess a singular part, and hence must be absolutely continuous. This argument immediately, when applied to a sequence of intervals $[0, b_n]$ and $[a_n, 1]$, leads to $\phi(0) = 0$, $\phi(1) = 1$.

We turn finally to showing that there cannot exist an interval $[a, b]$ with $b > a$ and $\phi(b) = \phi(a)$. Assume otherwise, and w.lo.g. let $a = 0$, and assume $\phi(b + \epsilon) > 0$ for all $\epsilon > 0$. Consider the curve $(x, \phi_\epsilon(x))$ which is linear between $(0, 0)$ and $(b + \epsilon, \phi(b + \epsilon))$: $\phi_\epsilon(x) = x\dot{\phi}_\epsilon, 0 \leq x \leq b + \epsilon$ with $\dot{\phi}_\epsilon = \frac{\phi(b+\epsilon)}{b+\epsilon}$ and $\phi_\epsilon(x) = \phi(x), b + \epsilon < x \leq 1$. Then

$$J(\phi_\epsilon) - J(\phi) = \int_0^{b+\epsilon} \sqrt{p(x, x\dot{\phi}_\epsilon)\dot{\phi}_\epsilon}\, dx - \int_b^{b+\epsilon} \sqrt{p(x,\phi)\dot{\phi}}\, dx \geq p\phi(b+\epsilon)^{1/2}(b+\epsilon)^{1/2} - P\phi(b+\epsilon)^{1/2}\epsilon^{1/2}.$$

Taking a small enough ϵ yields a contradiction, and completes the proof of the lemma. □

We return to the proof of the theorem. To see that the optimizing curve must satisfy the differential equation (24), we make use of the Hamiltonian form of the Pontryagin maximum principle. First, we check that all conditions needed to apply Theorem 5.1.i in [1] apply. This will lead to a version of (24) which holds almost everywhere. We then use the particular properties of the problem to guarantee that the optimizing path indeed satisfies (24).

To apply the above mentioned theorem of [1], let $f_o(t,x,u) = -\sqrt{p(t,x)u}$, $g=0$, $f(t,x,u)=u$, $B=\{0,0,1,1\}$, $A=[0,1]\times[0,1]$, $U=[0,\infty)$ and $M=A\times U$. Then, f_{ot} and f_{ox} are both continuous in M, conditions 4.1.a, 4.1.b, 4.1.d are easily checked to be satisfied, and condition 4.1.c' is satisfied since, for the optimizing path $x^*(t), u^*(t)$ (with $\dot{x}^*(t) = u^*(t)$ a.e.),

$$|f_{ot}(t',x',u^*(t))| \le c\sqrt{u^*(t)} \in L^1[0,1]$$

where the last inclusion is due to the fact that $u^* \ge 0$ and $\int_0^1 u^*(s)ds = 1$. A similar computation holds for f_{ox}. Thus, Theorem 5.1.i applies and, defining

$$H(t,x,u,\lambda) = -\sqrt{p(t,x)u} + \lambda u$$

and

$$M(t,x,\lambda) = \begin{cases} -\frac{p(t,x)}{4\lambda} & \lambda > 0 \\ -\infty & \lambda \le 0 \end{cases}$$

one concludes that, for some absolutely continuous $\lambda(t)$ satisfying the equation

$$\dot{\lambda}(t) = \frac{p_y(t,x^*(t))\sqrt{u^*(t)}}{2\sqrt{p(t,x^*(t))}} \text{ a.e.} \tag{25}$$

it holds that

$$-\sqrt{p(t,x^*(t))u^*(t)} + \lambda(t)u^*(t) = -\frac{p(t,x^*(t))}{4\lambda(t)} \text{ a.e..} \tag{26}$$

These equations imply that

$$2\sqrt{\lambda(t)u^*(t)} = \sqrt{p(t,x^*(t))/\lambda(t)} \text{ a.e..} \tag{27}$$

Let $\Theta = \{t : \lambda(t) = 0\}$. Since $p(x,y)$ is bounded away from zero, the Lebesgue measure of Θ is null, and the RHS of the last equation is continuous on $[0,1]\setminus\Theta$. Hence, $u^*(\cdot)$ is continuous there, and actually one may take $\sqrt{u^*(t)} = \sqrt{p(t,x^*(t))}/2\lambda(t)$ there. Moreover, the RHS of (27) is differentiable a.e., and since $\int_0^1 (\sqrt{u^*(t)})^2 dt = 1$, one deduces that $\int_0^1 \lambda(t)^{-2}dt < \infty$. One concludes that

$$\frac{\dot{u}^*(t)}{2\sqrt{u^*(t)}} = \frac{p_x(t,x^*(t)) + \dot{x}^*(t)p_y(t,x^*(t))}{4\lambda(t)\sqrt{p(t,x^*(t))}} - \frac{\dot{\lambda}(t)\sqrt{p(t,x^*(t))}}{2\lambda^2(t)}.$$

Recalling that $\dot{u} = \ddot{x} = \ddot{\phi}$, and using (25), one obtains (24).

Next, we show that $\dot{\phi}$ is bounded and $\dot{\phi}(\cdot) > 0$. Indeed, let $\psi = \log\dot{\phi}, q = \log p$, then, in terms of ψ and q, the equation (24) is

$$2q_x - \frac{d}{dx}q(x,\phi(x)) = \dot{\psi}.$$

Thus, for any $x_o \in (0,1)$,

$$\psi_o(x) - \psi(x_o) + q(x,\phi(x)) - q(x_o,\phi(x_o)) = 2\int_{x_o}^{x} q_x(s,\phi(s))\,ds$$

i.e.

$$\dot{\phi}(x) = \dot{\phi}(x_o)\frac{p(x_o, \phi(x_o))}{p(x, \phi(x))} \exp[2\int_{x_o}^{x} \frac{p_x}{p}(s, \phi(s))\, ds].$$

Now if $\sup_{x,y} p(x, y) = M, \inf_{x,y} p(x, y) = m$ and $\sup_{x,y} \frac{|p_x|}{p}(x, y) = N$, we get the bound

$$\dot{\phi}(x_o)\frac{p(x_o, \phi(x_o))}{M} \exp[-2N] \leq \dot{\phi}(x) \leq \dot{\phi}(x_o)\frac{p(x_o, \phi(x_o))}{m} \exp[2N].$$

Since $\int_0^1 \dot{\phi}(x)\, dx = 1$, we get

$$\frac{m}{p(0,0)} \exp[-2N] \leq \dot{\phi}(0) \leq \frac{M}{p(0,0)} \exp[2N].$$

Finally let us show the last statement of the theorem: Let ϕ_1 and ϕ_2 be two solutions of the variational problem and suppose that $\phi_1(x_0) = \phi_2(x_0)$ for some $x_0 \in (0, 1)$. Then $\phi(x) = \phi_1(x)$, $x \leq x_0$ and $\phi(x) = \phi_2(x)$ $x \geq x_0$, is again solution and solves the Euler equation (24). But this implies $\dot{\phi}(x_0) = \dot{\phi}_1(x_0) = \dot{\phi}_2(x_0)$ and by uniqueness of the Euler equation $\phi_1(x) = \phi_2(x), 0 \leq x \leq 1$.

□

The following is an immediate corollary:

Corollary 1 **(A1)–(A3)** *imply* **(A4)**.

Proof : Use the lower semicontinuity of $G(\phi)$ together with the continuity of the solutions of the variational problem which was proved in Theorem 4. □

4 Examples

In this section, we provide some examples of densities $p(x, y)$ which satisfy the assumptions **(A1), (A3)** and either **(A2), (A2')** or **(A2")**.

The uniform case

Let $p(x, y) = 1$. Obviously, $p(x, y)$ satisfies **(A1)-(A2)**. Moreover, we claim that the optimization problem (3) possesses the unique solution $\phi_1(x) = x$. Indeed, suppose there exists a $1 > \delta > 0$ such that $\phi(t) = t + \delta$ for some $t \in [0, 1 - \delta]$ (the symmetric case $\phi(t) = t - \delta$ works in the same way). Then, Jensen's inequality implies

$$\begin{aligned} \int_0^1 \dot{\phi}(x)^{1/2}\, dx &= \frac{1}{t}\int_0^t t\dot{\phi}(x)^{1/2}\, dx + \frac{1}{1-t}\int_t^1 (1-t)\dot{\phi}(x)^{1/2}\, dx \\ &\leq \left(t\int_0^t \dot{\phi}(x)\, dx\right)^{1/2} + \left((1-t)\int_t^1 \dot{\phi}(x)\, dx\right)^{1/2} \end{aligned}$$

$$
\begin{aligned}
&= t^{1/2}(t+\delta-\phi(0))^{1/2}+(1-t)^{1/2}(\phi(1)-t-\delta)^{1/2} \\
&\leq t^{1/2}(t+\delta)^{1/2}+(1-t)^{1/2}(1-t-\delta)^{1/2} \\
&\leq (1-\delta^2)^{1/2}\,,
\end{aligned}
\tag{28}
$$

where the maximum is achieved at $\phi(0)=0, \phi(1)=1$ and $t=\frac{1-\delta}{2}$. Since $J(\phi_1)=1$, this proves the claim.

The piecewise constant case

Let $\{S_{ij}\}$ be a finite partition of $[0,1]^2$ into squares and let $p(x,y)$ be a positive constant on each square S_{ij}. By the same Jensen inequality argument as in the uniform case, it follows that the minimizing curve (curves) are piecewise linear, with constant slope on each square S_{ij}. Moreover, the optimization problem (3) becomes then an optimization problem over the set of slopes. We shall prove that in this case **(A3)** is always satisfied.

More precisely let $S_{ij}=[i/N,(i+1)/N)\times[j/N,(j+1)/N)$ and assume that $p(i,j)>0$. We choose first an increasing collection of squares $\{S_{i_k,j_k}, k=1,...,2N-1\}$ crossed by the trajectory: with $i_1=j_1=0$, $i_{2N-1}=j_{2N-1}=N-1$ and $i_{k+1}\geq i_k, j_{k+1}\geq j_k$, $i_{k+1}-i_k+j_{k+1}-j_k=1$. An optimal trajectory is an increasing line with constant slope on each S_{i_k,j_k} passing through $\{(x_k,y_k): k=0,...,2N-1\}$ with $(x_0,y_0)=(0,0)$, $(x_{2N-1},y_{2N-1})=(1,1)$. From Lemma 6, we know that $w_k=x_k-x_{k-1}>0$ and $z_k=y_k-y_{k-1}>0, k=1,...,2N-1$. (Note that when proving that the number of optimizing curves is finite, we may assume that interior points $\{(x_k,y_k): k=1,...,2N-1\}$ do not fall on the corners of squares S_{i_k,j_k}, i.e., one, and only one, of x_k or y_k intersects the grid $\{(i_k/N,j_k/N)\}$, since allowing for path crossing the corners only adds a finite number of possibilities. See Figure 1). Write

$$
J(w,z)=\sum_{k=1}^{2N-1} p_k^{1/2}w_k^{1/2}z_k^{1/2},
$$

where p_k is the value of $p(i_k,j_k)$ in the square $S_{i_kj_k}$.

We are going to show that $J(w,z)$ is strictly concave on admissible (w,z). Thus for a given collection of squares $\{S_{i_k,j_k}\}$, there exists a unique non-degenerate maximizing curve. Since there are only finitely many possible choices, we see that both **(A3)** and **(A4)** are satisfied in the piecewise constant case.

To see the strict concavity of $J(w,z)$, note that if $\{(x_k,y_k)\}$ and $\{(x'_k,y'_k\}$ are admissible (i.e., positive, in $[0,1]^2$, and such that the corresponding $\{(w_k,z_k)\}$ and $\{(w'_k,z'_k)\}$ are strictly positive), so is their convex combination, and the latter leads to the same convex combination of the $\{(w_k,z_k)\}$ vectors. On the other hand, note (by computing the Hessian) that the function $\sqrt{xy}$ is concave, and strictly concave for pairs (x,y) and (x',y') such that $x/y\neq x'/y'$. Since for some $m\in\{1,\dots 2N-1\}$, it holds that either $w_m\neq w'_m$ or $z_m\neq z'_m$, and since for the smallest such m, $z_m/w_m\neq z'_m/w'_m$, the strict concavity follows.

To see an example where the solution to (4) is not unique, consider the density $p(x,y)=p_1$ for $(x,y)\in[0,1/2]^2\cup[1/2,1]^2$ and $p(x,y)=p_2$ otherwise. It is easy to check that the solution to

Figure 1: Piecewise constant example - $N = 2$

(3) is the diagonal $\phi(x) = x$ if $p_1 > p_2$, whereas if $p_2 > p_1$ then the following curves are the only maximizing curves:

$$\phi_1(x) = \begin{cases} p_2 x/p_1 & 0 \leq x \leq p_1/2p_2 \\ \frac{1}{2} + (x - \frac{p_1}{2p_2}) & p_1/2p_2 < x \leq 1/2 \\ y^* + (2x-1)(1-y^*) & 1/2 < x \leq 1 \end{cases}$$

where

$$y^* = 1 - \frac{p_1}{2p_2},$$

and, since by symmetry $\phi_2(x) = 1 - \phi_1(1-x)$,

$$\phi_2(x) = \begin{cases} p_1 x/p_2 & 0 \leq x \leq 1/2 \\ p_1/2p_2 + (x - 1/2) & 1/2 < x \leq y^* \\ \frac{1}{2} + \frac{x-y^*}{2(1-y^*)} & y^* < x \leq 1 \end{cases}$$

Note, in this set-up, that although the density is symmetric, the maximizing curve is not the diagonal!

Checkerboard

As another piecewise constant example, let $1 \leq i, j \leq 7$, let $S_{i,j} = [(i-1)/7, i/7) \times [(j-1)/7, j/7)$, and let

$$p(x,y) = \begin{cases} 49/9 & (x,y) \in S_{ij},\ (i,j) \in \{(1,1),(2,3),(3,2),(3,5),(4,4),(5,3),(5,6),(6,5),(7,7)\} \\ 0 & \text{otherwise} \end{cases}$$

(note that $p(x,y) > 0$ on all possible increasing paths of a knight, starting at $(1,1)$ and progressing to $(7,7)$, on a 7×7 checkerboard). It is easy to check that in this case, the optimal paths are all composed of diagonal segments of the S_{ij}'s where $p(x,y) > 0$, and every nondecreasing arrangement of such diagonal segments with $\phi(0) = 0$ and $\phi(1) = 1$ is optimal (the value of the path on squares with $p(x,y) = 0$ is of no importance since there are (a.s.) no samples in these squares. For consistency, it should be chosen such that the resulting path is in $B^{\nearrow}$).

Figure 2: Checkerboard example

Convex problems

Next, we consider the case of $p_{yy} \leq 0$. It is straightforward to check that in this case, $J(\phi)$ is strictly concave. It follows that the maximizing curve is unique. The problem being symmetric with respect to the x and y coordinates, it is obvious that the same conclusion may be drawn if $p_{xx} \leq 0$.

Independent coordinates

Although the case of independent coordinates, characterized by $p(x,y) = f(x)g(y) > 0$, may be reduced to the uniform case by a change of coordinate, it is interesting to note that in this case, the unique solution of (24) is provided by the equation $F(x) = G(\phi(x))$, where $F(x) = \int_0^x f(\theta)d\theta$ and $G(y) = \int_0^y g(\theta)d\theta$.

An explicit, continuous example

We describe below an example where $p(x,y)$ satisfy **(A1)**–**(A3)**, yet the optimal curves do not include the diagonal (although the diagonal is a solution of (24), it is not a maximizer). This example also illustrates the case of multiple maximizers.

Consider a density of the form

$$p(x,y) = k_F^2 \exp[F((x-y)^2)],$$

where $F \in C_b^1$ with $F(0)=0$ and $k_F > 0$ is a normalizing constant. In this case (24) becomes

$$\ddot{\phi} = 2F'((x-\phi)^2)(x-\phi)(\dot{\phi}+\dot{\phi}^2)$$

with boundary condition $\phi(0)=0, \phi(1)=1$. Note that $\phi_1(x)=x$ is a solution to the equation.

In case $F'<0$, ϕ_1 is the unique solution: namely suppose that $\phi(x)<x$ for some $x\in(0,1)$, then $\ddot{\phi}(x)<0$, i.e. the curve is concave and remains below the diagonal, i.e. $\phi(1)\neq 1$.

In case $F'>0$, we have two other solutions ϕ_2 and ϕ_3. ϕ_2 is strictly concave with $\dot{\phi}_2(0)>1$, $\phi_2(x)>x, x\in(0,1)$, and $\phi_3=\phi_2^{-1}$ is strictly convex with $\dot{\phi}_3(0)=\frac{1}{\dot{\phi}_2(0)}<1$ and $\phi_3(x)<x, x\in(0,1)$.

Let us show that for $F'(0)>3$, ϕ_1 cannot be optimal. For $0<\epsilon<\infty$ consider the piecewise linear curve $(x,\psi_\epsilon(x))$ passing through $(0,0)$, $(t(\epsilon),1-t(\epsilon))$ and $(1,1)$ where $t(\epsilon)=\frac{1}{2+\epsilon}$. Of course, $\psi_0=\phi_1$ and

$$J(\psi_\epsilon) = 2k_F(1+\epsilon)^{1/2}\int_0^{t(\epsilon)} \exp[F(\epsilon^2x^2)/2]\,dx.$$

This yields

$$\begin{aligned}\tfrac{d}{d\epsilon}J(\psi_\epsilon) &= \tfrac{1}{2}(1+\epsilon)^{-1}J(\psi_\epsilon) + 2k_F(1+\epsilon)^{1/2}t'(\epsilon)\exp[F(\epsilon^2 t(\epsilon)^2)/2] \\ &+2\epsilon k_F(1+\epsilon)^{1/2}\textstyle\int_0^{t(\epsilon)} \exp[F(\epsilon^2x^2)/2]F'(\epsilon^2x^2)x^2\,dx.\end{aligned}$$

At $\epsilon=0$ we get

$$\frac{d}{d\epsilon}J(\psi_\epsilon)\Big|_{\epsilon=0} = \frac{1}{2}J(\psi_0) - \frac{k_F}{2} = 0.$$

Next

$$\frac{d^2}{d\epsilon^2}J(\psi_\epsilon)\Big|_{\epsilon=0} = -\frac{1}{2}J(\psi_0) - \frac{k_F}{4} + \frac{k_F}{2} + 2k_F\int_0^{1/2}\exp[F(0)]F'(0)x^2\,dx = -\frac{k_F}{4}+\frac{k_F}{12}F'(0) > 0,$$

thus for sufficiently small $\epsilon>0$, $J(\psi_\epsilon)>J(\phi_1)$.

We believe that the condition $F'(0)>3$ for the existence of multiple solutions is redundant: indeed, in the case $F(z)=c\sqrt{z}$, $c>0$ (which unfortunately does not satisfy the smoothness assumption in (A2)), one can show that

$$dJ(\psi_\epsilon)/d\epsilon|_{\epsilon=0} > 0\,.$$

It follows that the off-diagonal solutions of (24), namely

$$\phi_2(x) = \int_0^x \frac{1}{(1+e^{-c})e^{ct}-1}dt = \frac{1}{c}\log\left((1+e^{-c})e^{cx}-1\right)+1-x,\ \ \phi_3(x)=\phi_2^{-1}(x)\,,$$

are optimal.

5 Longest increasing subsequence

In this section, we consider the length and location of the longest increasing subsequence of an i.i.d., two dimensional sample. We provide here the

Proof of Theorem 2: We begin by fixing some notations. Let $\ell_{\max}(n)$ denote the length of the longest increasing subsequence, and denote the corresponding increasing subsequence (which may not be unique) by $\mathcal{Z} = ((x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), \ldots, (x_{\ell_{\max}(n)}, y_{\ell_{\max}(n)}))$. A famous result of Vershik and Kerov (see [8, 9]) states that if $P(\cdot,\cdot)$ is the uniform measure on the unit square, then $\ell_{\max}(n)/\sqrt{n} \to_{n\to\infty} 2$ in probability.

As in the proof of Theorem 1, we will use upper and lower bounds on the probability of the longest increasing subsequence being around a given path $\phi(\cdot)$. Before doing that, we need the following lemma.

Lemma 7 *Assume $(1-\delta) < p(x,y) < (1+\delta)$. Then there exist $c_\delta > 0$, which depend on δ only (and not on the specific law $P(\cdot,\cdot)$), such that*

$$\lim_{n\to\infty} P(|\frac{\ell_{\max}(n)}{\sqrt{n}} - 2| > c_\delta) = 0\,,$$

and $\lim_{\delta\to 0} c_\delta = 0$.

Proof of Lemma 7 : Note that by a suitable change of coordinates in the x-axis, we may and will assume that $p(x) = \int p(x,y)dy = 1$. We use a coupling argument and the result of Vershik and Kerov. Let π be the (random) permutation such that $x_{\pi(1)} < x_{\pi(2)} < \cdots < x_{\pi(n)}$, let $P_i(\cdot)$ be the law on $[0,1]$ with density $p(y|x_{\pi(i)})$, and note that, for δ small enough, $1-\delta' < p(y|x_{\pi(i)}) < 1+\delta'$, with $\delta' = 2\delta/(1-\delta)$. To see first the lower bound in the statement of the lemma, note that P_i may be written as a mixture of a uniform law (with weight $(1-\delta')$) and another law on $[0,1]$, denoted q_i. Thus, the sample $(x_{\pi(1)}, y_{\pi(1)}), \ldots, (x_{\pi(n)}, y_{\pi(n)}))$ possesses the same law as $\tilde{Z}_n = (x_{\pi(1)}, (1_{m_1=1}U_1 + (1-1_{m_1=1})Z_1)), \ldots, (x_{\pi(n)}, (1_{m_n=1}U_n + (1-1_{m_n=1})Z_n))$, where $\{U_i\}_{i=1}^n$ is a sequence of i.i.d. uniform random variables, independent of the sequence $\{x_i\}_{i=1}^n$, $\{m_i\}_{i=1}^n$ is a sequence of i.i.d. Bernoulli$(1-\delta')$ random variables, independent of the sequences $\{U_i\}_{i=1}^n$ and $\{x_i\}_{i=1}^n$, and $\{Z_i\}_{i=1}^n$ is a sequence of random variables whose law depends on the sequence $\{x_i\}_{i=1}^n$ (explicitely, $P(Z_i \in dx) = (P_i(dx) - (1-\delta')1_{[0,1]}(x)dx)/\delta'$. Let I denote the set of indices with $m_i = 1$, and let $N_n = \sum_{i=1}^n 1_{m_i=1} = |I|$ denote the number of indices where a uniform random variable is chosen in the mixture. Note that $P(N_n/n < 1-2\delta') \to_{n\to\infty} 0$ due to the law of large numbers. Let $\tilde{\ell}_{\max}(n)$ denote the length of the maximal increasing subsequence corresponding to $\tilde{Z}_n$, then $\tilde{\ell}_{\max}(n)$ possesses the same law as $\ell_{\max}(n)$ and, on the other hand, is not smaller than the length of the maximal increasing subsequence when one considers only those indices $i \in I$. The latter is distributed precisely as the length of the maximal increasing subsequence of a *uniform* sample of random length N_n which is independent of the uniform sequence. Hence, by the result of Vershik and Kerov, $\lim_{n\to\infty} P(\tilde{\ell}_{\max}(n)/\sqrt{n} \le 2\sqrt{1-2\delta'}) = 0$, which concludes the proof of the lower bound.

The proof of the upper bound is similar. Indeed, let π be as before, and decompose now uniform random variables to a mixture of random variables distributed according to P_i (with weight $1/(1+\delta')$) and auxiliary random variables Z_i. The argument used in the proof of the lower bound applies now to this case with the obvious modifications.

$\square$

We now return to the proof of the theorem. We begin with the following lemma.

Lemma 8 *Let $\phi \in B^{\nearrow}$ be a C_b^1 curve. For any $\delta > 0$, define the event*

$$A_n = \{\omega : \exists \text{increasing subsequence of length} \geq 2(J(\phi)-\delta)\sqrt{n} \text{ wholly contained in a } \delta \text{ neighborhood of } \phi(\cdot)\}.$$

Then,

$$\lim_{n\to\infty} P(A_n) \to 1 .$$

Proof of Lemma 8 : Fix an integer K, and let $\Delta x = 1/K$. Let $Y_i = \phi(i\Delta x)$, and increase K if necessary such that $\max_i Y_{i+1} - Y_i < \delta$. Define the rectangles $R_i = [i\Delta x, (i+1)\Delta x) \times [Y_i, Y_{i+1})$, $i = 0, 1, \ldots, K-1$. Reduce K further if necessary such that, for δ' to be chosen below independently of K,

$$\max_i \max_{x,y\in R_i} \max(p(x,y)/p(i\Delta x, Y_i), p(i\Delta x, Y_i)/p(x,y)) < (1+\delta'),$$

and

$$\sum_{i=0}^{K-1} \sqrt{p(i\Delta x, \phi(i\Delta x))(\phi((i+1)\Delta x) - \phi(i\Delta x))\Delta x} \geq (1-\delta')J(\phi).$$

Let n_i denote the number of points of the sample $\{z_\alpha\}_{\alpha=1}^n$ in R_i, and note that by the law of large numbers,

$$\lim_{n\to\infty} P(\max_i \left| \frac{n_i}{np(i\Delta x, Y_i)[\phi((i+1)\Delta x) - \phi(i\Delta x)]\Delta x} - 1 \right| > 2\delta') = 0.$$

Note that (c.f. Lemma 2), the law of the sample conditioned on being in R_i, is i.i.d. with the density (on R_i) $p(x,y)/\int_{R_i} p(x,y)dxdy$. Hence, Lemma 7 may be applied and leads to the conclusion that

$$P(\text{There exists an increasing subsequence of length } 2(1-\delta'')\sqrt{n_i} \text{ in each of the rectangles } R_i) \to_{n\to\infty} 1,$$

where $\delta'' = \delta''(\delta') \to_{\delta'\to 0} 0$ independently of the value of K. Thus, choosing δ' small enough, with probability converging to 1 with n there exists an increasing subsequence in a δ neighborhood of ϕ, whose length is at least

$$\sum_{i=0}^{K-1} 2(1-\delta'')\sqrt{n}\sqrt{p(i\Delta x, \phi(i\Delta x))(\phi((i+1)\Delta x) - \phi(i\Delta x))\Delta x}\sqrt{1-2\delta'} \geq 2(1-\delta')(1-\delta'')\sqrt{n}J(\phi)/(1+2\delta').$$

The lemma follows by noting that δ' is arbitrary.

$\square$

Much as in the proof of Theorem 1, we need also an upper bound on maximal increasing subsequences. To this end, we introduce some definitions. Fix K a large integer as before, let β be a large integer and define $\Delta x = 1/K$, $\Delta y = \Delta x/\beta$. Let the multi-indices $\mathbf{t} = (t_1, \ldots, t_K)$ and $\mathbf{b} = (b_1, \ldots, b_K)$ be *admissible* if $b_1 \geq 0, b_K \leq 1/\Delta y$ and $t_1 \geq 0$, $b_i \leq t_i \leq 1/\Delta y - 1$, $i = 1, \ldots, K$ and $b_i \geq t_{i-1}$, $i = 2, \ldots, K$. Note that the number of admissible multi-indices is bounded above by $(1/\Delta y)^{2/\Delta x}$. We say that a multi-index $\mathbf{i} = (i_1, \ldots, i_{\ell_{\max}(n)})$ forms a $(\mathbf{t}, \mathbf{b})$ increasing subsequence if, for $(x_{i_1}, \ldots, x_{i_{\ell_{\max}(n)}})$ which is an increasing subsequence, the inclusion $x_{i_\ell} \in [i\Delta x, (i+1)\Delta x)$ implies that $y_{i_\ell} \in [b_i \Delta y, (t_i + 1)\Delta y)$. Note that being a $(\mathbf{t}, \mathbf{b})$ increasing subsequence depends on the sample $\{z_\alpha\}$, and that every $(\mathbf{t}, \mathbf{b})$ increasing subsequence defines a polygonal nondecreasing curve $\phi(\cdot) : [0,1] \to [0,1]$.

Finally, let

$$J_{(\mathbf{b},\mathbf{t})} = \sum_{i=1}^{K} \sqrt{p(i\Delta x, b_i \Delta y)(t_i - b_i + 1)\Delta x \Delta y}\,.$$

Note that $\lim_{\beta\to\infty} \lim_{\Delta x \to 0} \sup_{(\mathbf{b},\mathbf{t})} J_{(\mathbf{b},\mathbf{t})} \leq \bar{J}$.

Lemma 9 *For any $\delta > 0$ small enough, all Δx small enough, and all admissible $(\mathbf{b}, \mathbf{t})$ (which depend on Δx),*

$$P(\exists (\mathbf{b},\mathbf{t})\text{increasing sequence of length greater than } 2\sqrt{n}(J_{(\mathbf{b},\mathbf{t})} + \delta)) \to_{n\to\infty} 0\,.$$

Proof of Lemma 9 : Let $R_i = [i\Delta x, (i+1)\Delta x) \times [b_i \Delta y, (t_i+1)\Delta y)$, and let $\ell_i(n)$ denote the length of the maximal increasing subsequence wholly in R_i. Fix $\delta_1 > 0$, and define

$$\Theta_1((\mathbf{b},\mathbf{t})) = \{i : (t_i - b_i + 1)\Delta y \leq \delta_1\}\,,$$

$$\Theta_2((\mathbf{b},\mathbf{t})) = \{i : 1 \geq (t_i - b_i + 1)\Delta y > \delta_1\}\,.$$

Since, for any $(\mathbf{b}, \mathbf{t})$ increasing subsequence of length ℓ, $\ell \leq \sum_{i=1}^{K} \ell_i(n)$, one has

$$\begin{aligned} P(\exists (\mathbf{b},\mathbf{t})\text{increasing sequence of length greater than } 2\sqrt{n}(J_{(\mathbf{b},\mathbf{t})} + \delta)) \quad &\leq \\ P(\exists i \in \Theta_1((\mathbf{b},\mathbf{t}))\text{such that } \ell_i(n) \geq 2\sqrt{n}(\sqrt{p(i\Delta x, b_i\Delta y)(t_i - b_i + 1)\Delta x \Delta y} + \delta \Delta x/2)) \quad &+ \\ P(\sum_{i\in\Theta_2((\mathbf{b},\mathbf{t}))} \ell_i(n) > \delta\sqrt{n})\,. \qquad & \end{aligned} \tag{29}$$

Note that, for $\Delta y < \delta_1^2$, $|\Theta_2((\mathbf{b},\mathbf{t}))| < \delta_1^{-1} + 1$. Hence,

$$P(\sum_{i\in\Theta_2((\mathbf{b},\mathbf{t}))} \ell_i(n) > \delta\sqrt{n}) \leq (\delta^{-1} + 1) \max_{i\in\Theta_2((\mathbf{b},\mathbf{t}))} P(\ell_i(n) \geq \delta\delta_1\sqrt{n}/2)\,.$$

Let $\ell^U_{\max}(n)$ denote the length of the maximal increasing subsequence in an i.i.d. sample of length n of uniformly distributed random variables (on $[0,1]^2$). Since $c^{-1} < p(\cdot,\cdot) < c$ for some $c \geq 1$, we have, by the result of Vershik and Kerov and an argument similar to the one in Lemma 7, that, for any $i \in \Theta_2((\mathbf{b},\mathbf{t}))$, and n large enough,

$$P(\ell_i(n) > \delta\delta_1\sqrt{n}/2) \leq P(\ell^U_{\max}(k_c \Delta x n) > \delta\delta_1\sqrt{n}/2) \to_{n\to\infty} 0\,, \tag{30}$$

as soon as $\sqrt{k_c \Delta x} < \delta\delta_1$ (here, k_c depends on c only and is contributed by the fact that for large n, the number of sample points in R_i is certainly less than $2cn\Delta x$, while the conditional law in R_i is bounded above by c^2). On the other hand,

$$P(\exists i \in \Theta_1((\mathbf{b},\mathbf{t})) \text{such that } \ell_i(n) \geq 2\sqrt{n}(\sqrt{p(i\Delta x, b_i \Delta y)(t_i - b_i + 1)\Delta x \Delta y} + \delta \Delta x/2)) \quad \leq$$
$$(\Delta x)^{-1} \max_{i \in \Theta_1((\mathbf{b},\mathbf{t}))} P(\ell_i(n) \geq 2\sqrt{n}(\sqrt{p(i\Delta x, b_i \Delta y)(t_i - b_i + 1)\Delta x \Delta y} + \delta \Delta x/2)) .$$

Note that the law of (x_α, y_α), conditioned on $(x_\alpha, y_\alpha) \in R_i$, is "almost uniform" in the sense of Lemma 7. Hence, for δ_1 small enough (first) and then Δx small, one has by applying Lemma 7 that

$$P(\ell_i(n) \geq 2\sqrt{n}(\sqrt{p(i\Delta x, b_i \Delta y)(t_i - b_i + 1)\Delta x \Delta y} + \delta \Delta x/2)) \to_{n\to\infty} 0 . \tag{31}$$

Combining (30,31) yields the Lemma.

□

The proof of both parts of Theorem 2 follows from Lemmas 8 and 9 in exactly the same way that Theorem 1 followed from Lemma 1 and 4.

□

Acknowledgment We thank Charles Goldie for suggesting the problem and providing us with [5]. We thank Alex Ioffe for several discussions concerning the variational problem (3), Maurice Cochand for discussing with us his conjecture which led to the results in Section 5, Persi Diaconis for bringing [9] to our attention, and Hermann Rost for useful discussions concerning the latter. We also thank the referee for numerous suggestions and corrections which led to improvements and simplifications in the paper. In particular, the simple proof of strict concavity in the piecewise constant example in Section 4 is due to him.

Appendix

Proof of Lemma 2: By monotone class, it is clearly enough to show that for any disjoint sets $A_l \times B_m$, $l, m = 1, \ldots, K$, with A_l, B_m closed intervals in $[0,1]$, it holds that

$$E(\int f dL_n) = E(\int f d\overline{L}_n) ,$$

where $f(x,y) = \sum_{l,m} \alpha_{lm} 1_{A_l \times B_m}(x,y)$. Without loss of generality, one may assume $A_l \times B_m$ to form a partition of $[0,1]^2$ which refines the partition generated by $\Delta x_i \times \Delta y_j$. It follows that it is enough to show that, for any vector a_{lm} with integer entries satisfying $\sum_{lm} a_{lm} = n$, $P(n_{lm} = a_{lm} \ \forall l, m) = P(\overline{n}_{lm} = a_{lm} \ \forall l, m)$, where $n_{lm} = |\{\alpha : z_\alpha \in A_l \times B_m\}|$ and $\overline{n}_{lm} = |\{\alpha : \overline{z}_\alpha \in A_l \times B_m\}|$. Note however that by a simple combinatorial computation,

$$P(n_{lm} = a_{lm} \ \forall l, m) = \prod_{l,m} P_{lm}^{a_{lm}} \frac{n!}{\prod_{l,m} a_{lm}!} ,$$

where $P_{lm} = \int_{A_l \times B_m} P(dz)$. On the other hand, let $\mathcal{A}_{ij} = \{(l,m) : A_l \times B_m \in \Delta x_i \times \Delta y_j\}$. Now,

$$P(\overline{n}_{lm} = a_{lm} \ \forall l,m) = \sum_{\mu \text{ is a type}} P(\overline{\ell}_n = \mu) P_\mu(\overline{n}_{lm} = a_{lm} \ \forall l,m)$$

Let $q_{ij} = \sum_{(l,m) \in \mathcal{A}_{ij}} a_{lm}$, and let $\hat{P}_{ij} = \int_{\Delta x_i \times \Delta x_j} P(dz)$, $\bar{P}_{ij}(l,m) = \frac{\int_{A_l \times B_m} dP(dz)}{\hat{P}_{ij}}$.

Let $Q(i,j,q_{ij})$ denote the probability that, out of nq_{ij} independent drawings on $\Delta x_i \times \Delta y_j$, each distributed according to $\bar{P}_{ij}$ on $\Delta x_i \times \Delta y_j$, exactly na_{lm} belong to each $A_l \times B_m$, $(l,m) \in \mathcal{A}_{ij}$. It follows that

$$\begin{aligned} P(\overline{n}_{lm} = a_{lm} \ \forall l,m) &= P(\overline{\ell}_n(i,j) = q_{ij} \ \forall i,j)(\prod_{i,j} Q(i,j,q_{ij})) \\ &= (\prod_{i,j} \hat{P}_{ij}^{q_{ij}})(\frac{n!}{\prod_{i,j} q_{ij}!}) \prod_{i,j}(\prod_{(l,m)\in\mathcal{A}_{ij}} \bar{P}_{ij}(l,m)^{a_{lm}})(\frac{q_{ij}!}{\prod_{(l,m)\in\mathcal{A}_{ij}} a_{lm}!}) \\ &= \prod_{l,m} P_{lm}^{a_{lm}} \frac{n!}{\prod_{l,m} a_{lm}!}. \end{aligned}$$

□

References

[1] L. Cesari, Optimization- Theory and applications. Springer, 1983.

[2] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Boston: Jones and Bartlett, 1992.

[3] I. Ekeland, Convexity methods in Hamiltonian mechanics, Springer, 1990.

[4] C. M. Goldie and S. I. Resnick, *Records in a partially ordered set*, Ann. Probab., 17 (1989), pp. 678–699.

[5] C. M. Goldie and S. I. Resnick, *Multivariate records and ordered random scattering*, preprint, 1993.

[6] A. D. Ioffe and V. M. Tichomirov, Theory of Extremal Problems. North-Holland, 1979.

[7] D. Pollard, Convergence of stochastic processes. Springer 1984.

[8] B.F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Advances in Math., 26 (1977) pp. 206–222.

[9] A. M. Vershik and C. V. Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables*, Dokl. Acad. Nauk. 233 (1977), pp. 1024–1028.

J.D. Deuschel
Department of Mathematics
Strasse des 17 Juni 135
Technical University, Berlin

O. Zeitouni
Department of Electrical Engineering
Technion
Haifa 32000, Israel