Many Multivariate Records

Charles M. Goldie[1] and Sidney I. Resnick[2]

Queen Mary & Westfield College (University of London) and Cornell University

January 2, 2000

Abstract

Suppose $(\mathbf{X}_n)_{n=1,2,\ldots}$ is a sequence of independent, identically distributed random vectors in $\mathbb{R}^2$. Say $\mathbf{X}_n$ is a record if there is a record simultaneously in both coordinates at index n. The total number of records of a sequence is frequently finite so we consider the behaviour of the records in a fixed rectangle A conditional on there being a large number of records N_A in the rectangle. We join the records up to form a random parametrized curve, and prove that, under suitable conditions, as the number of records in the rectangle goes to infinity the random curves converge conditionally in probability to a non-random limit parametrized curve, which we characterize. We also prove a large-deviation result for the random curves.

Keywords: Extreme values, Hazard measure, Large deviations, Limiting curve, Multivariate record values, Order statistics

AMS 1991 mathematics subject classification: 60F10, 60G70

1. Introduction.

Let $\mathbf{X}_n := (X_n^{(1)}, X_n^{(2)})$, for $n = 1, 2, \ldots$, be independent identically distributed (iid) $\mathbb{R}^2$-valued random vectors with common distribution F. We are interested in the records of such a sequence. For $\mathbb{R}$-valued random variables, it is clear what the definition of record has to be but in higher dimensions that is no longer the case. Many definitions are possible, some with applied appeal and some with mathematical significance. However the following definition, based on the natural partial ordering of $\mathbb{R}^2$, has allowed the most mathematical progress: say $\mathbf{X}_n$ is a record if there is a record simultaneously in both coordinates, that is, $X_n^{(1)} > \bigvee_{j=1}^{n-1} X_j^{(1)}$ and $X_n^{(2)} > \bigvee_{j=1}^{n-1} X_j^{(2)}$. For fuller discussion of this and other notions of multivariate records, see [Goldie & Resnick 1989], [Gnedin 1993] and [Goldie & Resnick 1995].

We define relations and operations among vectors component-wise. For example,

$$\mathbf{x} \vee \mathbf{y} := (x^{(1)} \vee y^{(1)}, x^{(2)} \vee y^{(2)}),$$
$$\mathbf{x} < \mathbf{y} \quad \text{means} \quad x^{(1)} < y^{(1)},\ x^{(2)} < y^{(2)},$$

and so on. Thus, in this notation, $\mathbf{X}_n$ is a record if $\mathbf{X}_n > \bigvee_{i=1}^{n-1} \mathbf{X}_i$.

Now consider records falling in a given closed rectangle A with sides parallel to the coordinate axes. Let $\mathbf{X}_0 := -\infty$ and $I_0^A := 0$. For $n = 1, 2, \ldots$ the indices when

[1]A visit to Cornell University was supported by Cornell's Mathematical Sciences Institute. This paper was revised during a visit to the University of Melbourne, supported by an SERC Block Travel Grant to Australia. Facilities provided by Cornell University's School of Operations Research and Industrial Engineering, and by the University of Melbourne's Department of Statistics, are gratefully acknowledged.

[2]Partially supported by NSF Grant DMS–9400535 at Cornell University.

records occur are

$$I_n^A := \inf\Big\{ m > I_{n-1}^A : \mathbf{X}_m \in A,\ \mathbf{X}_m > \bigvee_{j=1}^{m-1} \mathbf{X}_j \Big\}.$$

Here $\inf \emptyset := \infty$. For $n = 1, 2, \ldots$ the sequence of record values in A is

$$\mathbf{R}_n^A := \begin{cases} +\infty, & \text{if } I_n^A = \infty, \\ \mathbf{X}_{I_n^A}, & \text{if } I_n^A < \infty. \end{cases}$$

The total number of records in A is $N_A := \sup\{n : I_n^A < \infty\}$. From now on we omit the superscript A in $\mathbf{R}_n^A$.

Many properties of records are governed by the *hazard measure* H defined by

$$H(d\mathbf{x}) := \frac{F(d\mathbf{x})}{1 - F(\mathbf{x}-)} = \frac{P(\mathbf{X}_1 \in d\mathbf{x})}{P(\{\mathbf{X}_1 < \mathbf{x}\}^c)}.$$

For instance, according to [Goldie & Resnick 1989, Theorem 2.1], $P(N_A < \infty) = 1$ or $P(N_A = \infty) = 1$ according as $H(A) < \infty$ or $H(A) = \infty$. In fact, it is typical that $N_{\mathbb{R}^2} < \infty$, i.e., that there is only a finite number of records; see [Goldie & Resnick 1989, Theorem 2.1] and [Gnedin 1994a]. That is the case, for example, when F is any bivariate normal whose correlation is not $+1$ [Gnedin 1994b]. Although the sequence of record values is known to be a Markov process [Goldie & Resnick 1989, Theorem 2.6], little else is known about it and that coupled with the fact that the total number of records is typically finite led us to wonder what is the behaviour of the record sequence if we force a large number of records to occur in a compact box A. We prove that under fairly general sufficient conditions a parametrized curve based on the records converges in probability, conditionally on $N_A = n$, as $n \to \infty$, to a non-random limit curve.

We are also interested in how many records occur in A when $H(A) < \infty$. The random variable N_A is then known to be sub-Poisson [Goldie & Resnick 1989, §2] in the sense that $P(N_A = j) \le cP(N = j)$ for a constant c, where N is Poisson with parameter $H(A)$. Consequently $P(N_A > j) \le cP(N > j)$. Is the tail of N_A Poisson-like, as for the numbers of records falling in a chosen region in $\mathbb{R}$, or asymptotically smaller? N_A is inextricably bound up with where the records fall in A when many of them do so, so we must find out about the latter to get information about the former.

In §2 we show how the problem of convergence to the non-random limit curve may be reduced to where a large number of iid A–valued random variables $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ fall, given that they are ordered (Theorem 2.2). The common distribution of the $\mathbf{Z}_i$ is proportional to the hazard measure obtained from the common distribution of the $\mathbf{X}_i$. Thus it will not be a product distribution, even if the distribution of the $\mathbf{X}_i$ is.

An outline of the paper and its other main results is as follows. In §3 we prove (Theorem 3.2) that random parametrized curves based on the $\mathbf{Z}_j$ converge to a non-random limit, and in §4 (Theorem 4.1) we characterize the limit as the solution to a variational problem, and prove various properties of it. In §5 we establish the large-deviation behaviour of the random curves, and evaluate the rate of decay of the probability that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are totally ordered (all in Theorem 5.1). §6 is then devoted (Theorems 6.1, 6.3) to the application of the foregoing to the case of records, which is straightforward. We find that the tail of N_A is strictly sub-Poisson; in fact it is roughly like the probability that two independent Poisson random variables are both large.

Lemma 3.11, in §3, is also of independent interest. Given non-negative numbers $p_1, \ldots, p_k$, with sum 1, which have logarithms that are not allowed to vary by too much, it gives sharp upper and lower bounds on any partial sum of the p_j.

In each of §§2–6 the proofs and associated lemmas are held over to the end of the section. So for an overview of the results of the paper discussed so far one can just read the conclusions and discussion at the start of each of §§2–6.

A parallel and independent treatment of the problem is in [Deuschel & Zeitouni 1994]. Their results and ours complement each other very satisfactorily, as follows.

(a) They allow non-unique limit curves, while we provide a natural sufficient condition for a unique limit.
(b) They work with non-parametrized curves, and their limits are therefore the traces of ours on the set A. Our parametrized limits provide extra information about what proportion of the underlying observations gives rise to each portion of the limit-trace.
(c) They explore the variational problem that the limit curves solve, and exhibit a number of examples.
(d) We explore large deviations.
(e) We apply the various theorems for ordered random variables (the $\mathbf{Z}_j$) to records (the $\mathbf{X}_j$).
(f) They apply their results to the problem of the longest increasing subsequence of iid points in the plane, gaining thereby a very interesting generalization of the celebrated result of Vershik and Kerov on the longest increasing subsequence in a random permutation.

In the latter half of §4 we discuss the relation between Deuschel and Zeitouni's non-parametrized limit curve and ours. Some observations about our condition for uniqueness of the limit curve are in Remark 3.9.

We denote the Borel σ-algebra on a topological space S by $\mathcal{B}(S)$.

2. Reduction to ordered iid rvs.

Fix finite $\mathbf{a} < \mathbf{b}$ in $\mathbb{R}^2$ and set $A = [\mathbf{a}, \mathbf{b}] := \{\mathbf{x} : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$. Define the *avoidance function* Q by

$$Q(B) := P(N_B = 0) \qquad (B \in \mathcal{B}(\mathbb{R}^2)),$$

the probability of no records in B. It is characterized in Theorem 2.4 of [Goldie & Resnick 1989]. Denote by

$$A^{(n)} := \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in A^n : \mathbf{x}_1 < \cdots < \mathbf{x}_n\} \tag{2.1}$$

the set of all strictly increasing sequences of length n in A, and temporarily set $\mathbf{x}_0 := \mathbf{a}$, $\mathbf{x}_{n+1} := \mathbf{b}$. Then in terms of Q and H the joint distribution of the records in A is

$$P(\mathbf{R}_1 \in d\mathbf{x}_1, \ldots, \mathbf{R}_n \in d\mathbf{x}_n,\ N_A = n)$$
$$= H(d\mathbf{x}_1) \cdots H(d\mathbf{x}_n) \prod_{i=0}^{n} Q((\mathbf{x}_i, \mathbf{x}_{i+1})) \qquad ((\mathbf{x}_1, \ldots, \mathbf{x}_n) \in A^{(n)}) \tag{2.2}$$

[Goldie & Resnick 1989, Theorem 2.6], where $(\mathbf{a}, \mathbf{b})$ denotes the open rectangle $\{\mathbf{x} : \mathbf{a} < \mathbf{x} < \mathbf{b}\}$. It follows that

$$P(N_A = n) = \int_{A^{(n)}} H(d\mathbf{z}_1) \cdots H(d\mathbf{z}_n) \prod_{i=0}^{n} Q((\mathbf{z}_i, \mathbf{z}_{i+1})).$$

Now although Q can be calculated as an infinite series, it is difficult to compute explicitly in closed form and thus it is difficult to work with (2.2) directly. Fortunately however, we can neglect Q in the formulae for the limit curve, as we shall shortly prove.

Define $\lceil x \rceil := \min\{n \in \mathbb{N} : n \geq x\}$ for $x \geq 0$. The most straightforward way to present the record values in A as a parametrized curve is to consider $\mathbf{R}_{\lceil N_A \cdot \rceil} := (\mathbf{R}_{\lceil pN_A \rceil})_{0 \leq p \leq 1}$, with the convention $\mathbf{R}_0 := \mathbf{a}$. This is a random element of D_L, the space of non-decreasing functions $\mathbf{g} : [0,1] \to \mathbb{R}^2$ that are left-continuous on (0,1] and have $\mathbf{g}(0) = \mathbf{a}$ and $\mathbf{g}(1) \leq \mathbf{b}$. On D_L we shall need a uniform metric

$$\rho_u(\mathbf{f}, \mathbf{g}) := \sup_{0 \leq p \leq 1} |\mathbf{f}(p) - \mathbf{g}(p)| \qquad (\mathbf{f}, \mathbf{g} \in D_L),$$

where $|\cdot|$ is the maximum norm in $\mathbb{R}^2$: $|\mathbf{x}| = |(x^{(1)}, x^{(2)})| := |x^{(1)}| \vee |x^{(2)}|$.

Our final convergence results, Corollary 4.3 and Theorem 6.1, will be given in terms of the above space, but for intermediate working the most natural parametrization turns out to be in terms of inverses, as follows.

Write $\mathbf{a} = (a^{(1)}, a^{(2)})$, $\mathbf{b} = (b^{(1)}, b^{(2)})$ and for $i = 1, 2$ let $D_R^{(i)}$ be the space of right-continuous proper distribution functions concentrating on $[a^{(i)}, b^{(i)}]$. Our main space of parametrized paths will be $D_R := D_R^{(1)} \times D_R^{(2)}$. On D_R define a 'Lévy' metric

$$\rho_L(\mathbf{f}, \mathbf{g}) := \max(\rho_L^{(1)}(f^{(1)}, g^{(1)}), \rho_L^{(2)}(f^{(2)}, g^{(2)})),$$

where $\mathbf{f} = (f^{(1)}, f^{(2)})$, $\mathbf{g} = (g^{(1)}, g^{(2)})$, and

$$\rho_L^{(i)}(f^{(i)}, g^{(i)}) :=$$
$$\inf\{\varepsilon > 0 : f^{(i)}(x - \varepsilon) - \varepsilon \leq g^{(i)}(x) \leq f^{(i)}(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\}.$$

The metric space (D_R, ρ_L) is then compact. The Borel σ-algebra $\mathcal{B}(D_R)$ is that generated by the ρ_L topology. We also note that D_R is a convex set in that probability mixtures $p\mathbf{f} + (1-p)\mathbf{g}$ of elements of D_R remain elements of D_R. Hence it will make sense for functionals on this space to be described as concave, linear, etc.

We may embed D_R into the space M_A^1 of probability measures on A. Thus for any $\mathbf{f} = (f^{(1)}, f^{(2)}) \in D_R$ let $\mu(\mathbf{f}) := \mu^{(1)}(f^{(1)}) \times \mu^{(2)}(f^{(2)})$ where $\mu^{(i)}(f^{(i)})$ is the measure on $[a^{(i)}, b^{(i)}]$ determined by

$$\mu^{(i)}(f^{(i)})[a^{(i)}, x] := f^{(i)}(x) \qquad (a^{(i)} \le x \le b^{(i)}).$$

Then ρ_L–convergence in D_R is equivalent to narrow convergence of the corresponding product measures in M_A^1.

We shall also need a uniform metric on D_R, namely

$$\rho_u(\mathbf{f}, \mathbf{g}) := \max(\rho_u^{(1)}(f^{(1)}, g^{(1)}), \rho_u^{(2)}(f^{(2)}, g^{(2)})),$$

where $\rho_u^{(1)}$ and $\rho_u^{(2)}$ are the uniform metrics on $D_R^{(1)}$ and $D_R^{(2)}$ respectively. Recall that if $\mathbf{f}$ belongs to the set C of continuous elements of D_R, then ρ_u–convergence to $\mathbf{f}$ is equivalent to ρ_L–convergence to $\mathbf{f}$.

Given a finite non-empty collection $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of points in A we define $\mathbf{e}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in D_R$ by $\mathbf{e}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (e^{(1)}, e^{(2)})$ where $e^{(1)}$ and $e^{(2)}$ are the empirical distribution functions of the 1 and 2 components of the $\mathbf{x}_j = (x_j^{(1)}, x_j^{(2)})$:

$$e^{(i)}(t) := \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{x_j^{(i)} \le t\} \qquad (i = 1, 2;\ t \in \mathbb{R}).$$

The parametrized path joining the records is defined as $\mathbf{L} := \mathbf{e}(\mathbf{R}_1, \ldots, \mathbf{R}_{N_A})$, if $N_A > 0$. If $N_A = 0$ put $\mathbf{L} := (\mathbf{1}_{[a^{(1)}, \infty)}, \mathbf{1}_{[a^{(2)}, \infty)})$, say. So $\mathbf{L}$ is a random element of D_R.

Remark 2.1. Taking a component-wise inverse of a parametrized path as defined here retrieves our first notion of parametrized path. Thus for $\mathbf{f} = (f^{(1)}, f^{(2)}) \in D_R$ define $\mathbf{f}^{\leftarrow} = (f^{(1)\leftarrow}, f^{(2)\leftarrow})$ by

$$f^{(i)\leftarrow}(p) := \inf\{t : a^{(i)} \le t \le b^{(i)},\ f^{(i)}(t) \ge p\} \qquad (i = 1, 2;\ 0 \le p \le 1). \tag{2.3}$$

Then $\mathbf{f}^{\leftarrow} \in D_L$. Further, one may check that, on the event $N_A > 0$, $\mathbf{L}^{\leftarrow}$ is $\mathbf{R}_{\lceil N_A \cdot \rceil}$.

THEOREM 2.2. *Assume $0 < H(A) < \infty$ and that H has bounded density h on A. Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be iid random vectors with common distribution*

$$P(\mathbf{Z}_1 \in d\mathbf{z}) = \frac{H(d\mathbf{z})}{H(A)} \mathbf{1}_A(\mathbf{z}). \tag{2.4}$$

Let $Z_{n1}^{(i)} \le \cdots \le Z_{nn}^{(i)}$ be the $Z_1^{(i)}, \ldots, Z_n^{(i)}$ in increasing order, and set $Z_{n0}^{(i)} := a^{(i)}$, $Z_{n,n+1}^{(i)} := b^{(i)}$. Suppose that, for each $\varepsilon > 0$,

$$P\Big(\bigvee_{j=0}^{n} (Z_{n,j+1}^{(1)} - Z_{nj}^{(1)}) > \varepsilon \;\Big|\; \mathbf{Z}_1 < \cdots < \mathbf{Z}_n\Big) \to 0 \qquad (n \to \infty). \tag{2.5}$$

Define $\mathbf{M}_n := \mathbf{e}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$*, and on* D_R *define probability laws* μ_n*,* ν_n *by*

$$\mu_n(B) := P(\mathbf{M}_n \in B \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n),$$
$$\nu_n(B) := P(\mathbf{L} \in B \mid N_A = n) \qquad (B \in \mathcal{B}(D_R)).$$

Then $\|\mu_n - \nu_n\|_{\mathrm{var}} \to 0$ *as* $n \to \infty$*, where* $\|\cdot\|_{\mathrm{var}}$ *is total-variation norm (cf. e.g. [Deuschel & Stroock 1989, p. 64]). In particular, existence of a non-random* $\mathbf{m} \in D_R$ *such that for all* $\varepsilon > 0$*,*

$$P(\rho_L(\mathbf{M}_n, \mathbf{m}) > \varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \to 0 \qquad (n \to \infty) \tag{2.6}$$

implies that also, for all $\varepsilon > 0$*,*

$$P(\rho_L(\mathbf{L}, \mathbf{m}) > \varepsilon \mid N_A = n) \to 0 \qquad (n \to \infty). \tag{2.7}$$

Remark 2.3. (2.6) says that conditional on the $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ arranging themselves in increasing order, $\mathbf{M}_n$ converges in probability to a non-random limit. The measure H determines the distribution of the $\mathbf{Z}_j$ as in (2.4), and we shall prove in §3 for a fairly general class of measures H, not only hazard measures, that the $\mathbf{Z}_j$ satisfy (2.6). This has some independent interest. We will also establish enough continuity properties of the limit $\mathbf{m}$ that the convergence (2.6) will imply (2.5). Then, when in §6 we specialize H to hazard measure, Theorem 2.2 will give us (2.7).

In the proof of our large-deviation result for records, Theorem 6.3, we will show that the probabilities $\mu_n(B)$ and $\nu_n(B)$ become close in a different (ratio) sense from that of total-variation norm. (2.7) could alternatively, but less directly, be deduced from Theorem 6.3.

Some more notation: λ denotes Lebesgue measure on R^2. For any $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f^{\vee}(B) := \sup_{x \in B} f(x), \ \ f^{\wedge}(B) := \inf_{x \in B} f(x).$$

Proof of Theorem 2.2: Set $\mathbf{Z}_{nj} := (Z_{nj}^{(1)}, Z_{nj}^{(2)})$ for $j = 0, \ldots, n+1$. Since

$$\begin{aligned}\sum_{j=0}^{n} H((\mathbf{Z}_{nj}, \mathbf{Z}_{n,j+1})) &\le h^{\vee}(A) \sum_{j=0}^{n} \lambda((\mathbf{Z}_{nj}, \mathbf{Z}_{n,j+1})) \\ &= h^{\vee}(A) \sum_{j=0}^{n} (Z_{n,j+1}^{(1)} - Z_{nj}^{(1)})(Z_{n,j+1}^{(2)} - Z_{nj}^{(2)}) \\ &\le h^{\vee}(A)(b^{(2)} - a^{(2)}) \bigvee_{j=0}^{n} (Z_{n,j+1}^{(1)} - Z_{nj}^{(1)}),\end{aligned}$$

we find from (2.5) that for every $\eta > 0$,

$$P\Big(\sum_{j=0}^{n} H((\mathbf{Z}_{nj}, \mathbf{Z}_{n,j+1})) > \eta \;\Big|\; \mathbf{Z}_1 < \cdots < \mathbf{Z}_n\Big) \to 0 \qquad (n \to \infty).$$

Therefore, for every $0 < \eta < 1$,

$$P\Big(\prod_{j=0}^{n}(1 - H((\mathbf{Z}_{nj}, \mathbf{Z}_{n,j+1}))) > 1 - \eta \;\Big|\; \mathbf{Z}_1 < \cdots < \mathbf{Z}_n\Big) \to 1.$$

Now for any Borel set B in $\mathbb{R}^2$,

$$H(B) = EN_B \geq P(N_B \geq 1) = 1 - Q(B),$$

so

$$1 \geq Q(B) \geq 1 - H(B), \tag{2.8}$$

whence, for all $n \geq n_0(\eta)$,

$$P\Big(\prod_{j=0}^{n} Q((\mathbf{Z}_{nj}, \mathbf{Z}_{n,j+1})) > 1 - \eta \;\Big|\; \mathbf{Z}_1 < \cdots < \mathbf{Z}_n\Big) \geq 1 - \eta.$$

We have for $\mathbf{a} < \mathbf{x}_1 < \cdots < \mathbf{x}_n < \mathbf{b}$ that (recalling (2.1) for the notation $A^{(n)}$)

$$\begin{aligned} P(\mathbf{Z}_j \in d\mathbf{x}_i \text{ for } j = 1, \ldots, n \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) &= \frac{\prod_{j=1}^{n} \frac{H(d\mathbf{x}_j)}{H(A)}}{\int_{A^{(n)}} \prod_{j=1}^{n} \frac{H(d\mathbf{z}_j)}{H(A)}} \\ &= \frac{\prod_{j=1}^{n} H(d\mathbf{x}_j)}{\int_{A^{(n)}} \prod_{j=1}^{n} H(d\mathbf{z}_j)}. \end{aligned} \tag{2.9}$$

Put $\mathbf{z}_0 := \mathbf{a}$, $\mathbf{z}_{n+1} := \mathbf{b}$. To bring the next calculation down in size let us write $\mathbf{e}(\mathbf{z})$ for $\mathbf{e}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$, $H(d\mathbf{z})$ for $H(d\mathbf{z}_1) \cdots H(d\mathbf{z}_n)$, and $Q(\mathbf{z})$ for $\prod_{j=0}^{n} Q((\mathbf{z}_j, \mathbf{z}_{j+1}))$. Then for every $B \in \mathcal{B}(D_R)$,

$$\begin{aligned} \nu_n(B) &= \frac{\int_{A^{(n)}} \mathbf{1}\{\mathbf{e}(\mathbf{z}) \in B\} Q(\mathbf{z})\, H(d\mathbf{z})}{\int_{A^{(n)}} Q(\mathbf{z})\, H(d\mathbf{z})} \\ &\geq (1-\eta) \frac{\int_{A^{(n)}} \mathbf{1}\{\mathbf{e}(\mathbf{z}) \in B\} \mathbf{1}\{Q(\mathbf{z}) > 1 - \eta\}\, H(d\mathbf{z})}{\int_{A^{(n)}} H(d\mathbf{z})} \\ &\geq (1-\eta) \left(\mu_n(B) - \frac{\int_{A^{(n)}} \mathbf{1}\{Q(\mathbf{z}) \leq 1 - \eta\}\, H(d\mathbf{z})}{\int_{A^{(n)}} H(d\mathbf{z})}\right) \\ &= (1-\eta)\Big(\mu_n(B) - P\Big(\prod_{j=0}^{n} Q((\mathbf{Z}_{nj}, \mathbf{Z}_{n,j+1})) \leq 1 - \eta \;\Big|\; \mathbf{Z}_1 < \cdots < \mathbf{Z}_n\Big)\Big). \end{aligned}$$

When $n \geq n_0(\eta)$ the conditional probability in the last expression is at most η, so the expression as a whole is at least $\mu_n(B) - 2\eta$. Thus, for these n,

$$\inf_{B \in \mathcal{B}(D_R)} (\nu_n(B) - \mu_n(B)) \geq -2\eta,$$

so that $\|\nu_n - \mu_n\|_{\text{var}} \leq 4\eta$. Since η was arbitrary, the result follows. $\square$

3. Convergence to a non-random limit.

A real-valued function g on A is called *L-superadditive* or *lattice-superadditive* if $g(y^{(1)}, y^{(2)}) - g(x^{(1)}, y^{(2)}) - g(y^{(1)}, x^{(2)}) + g(x^{(1)}, x^{(2)}) \geq 0$ for all $\mathbf{x} < \mathbf{y}$ in A. This property has the following characterization (see [Marshall & Olkin 1979, pp. 146, 150] and references therein).

Proposition 3.1. *Suppose $g : A \to \mathbb{R}$ is continuous on A and the mixed partial derivative $\partial^2 g(x^{(1)}, x^{(2)})/\partial x^{(1)} \partial x^{(2)}$ exists on A°. Then g is L-superadditive on A if and only if $\partial^2 g(x^{(1)}, x^{(2)})/\partial x^{(1)} \partial x^{(2)} \geq 0$ on A°.*

The substantive result of this section is as follows.

Theorem 3.2. *Suppose H is a measure on $\mathbb{R}^2$ that has a continuous density h on A with $\log h$ bounded and L-superadditive on A. Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be iid random vectors with common distribution given by (2.4). Then there exists a non-random $\mathbf{m} \in D_R$ such that for each $\varepsilon > 0$,*

$$P(\rho_L(\mathbf{M}_n, \mathbf{m}) > \varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \to 0 \qquad (n \to \infty). \tag{3.1}$$

Remark 3.3. We may change the conditioning to be that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are totally ordered. Thus under the same conditions as in the Theorem we easily gain the conclusion that there exists a non-random $\mathbf{m} \in D_R$ such that

$$P(\rho_L(\mathbf{M}_n, \mathbf{m}) > \varepsilon \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_n \text{ can be ordered}) \to 0 \qquad (n \to \infty) \tag{3.2}$$

for each $\varepsilon > 0$. For, neglecting the event of probability 0 where two or more of the $\mathbf{Z}_i$ coincide in some component, the conditioning event is the union of the disjoint events $\{\mathbf{Z}_{\pi(1)} < \mathbf{Z}_{\pi(2)} < \cdots < \mathbf{Z}_{\pi(n)}\}$ as π ranges over all permutations of $1, \ldots, n$. Within each sub-event the set-up is identical to that in the Theorem, apart from a re-labelling, so the extension to (3.2) is clear.

Remark 3.4. The proof of Theorem 3.2 proceeds by a sequence of lemmas. We work with a discretized version of the path $\mathbf{M}_n$, and our main estimate of its probability distribution in path-space, within the event that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are in increasing order, is Lemma 3.6, where upper and lower bounds are derived. These involve functionals $\underline{S}_k(\mathbf{p})$ and $\overline{S}_k(\mathbf{p})$ of the discretized path $\mathbf{p}$, and a product of factorials. The latter we eventually evaluate by Stirling's formula (in the final stages of the proof of Theorem 3.2), and it will ultimately give rise to an entropy contribution to the large-deviation rate-function. But first a lot of work is needed to prove concavity properties of the functionals $\underline{S}_k$ and $\overline{S}_k$, sufficient to establish that the distribution of the discretized path *concentrates* on its most likely location. That identifies the limit path, to within the mesh of the discretization. A final compactness argument then gives the actual path convergence.

Proofs.

Pick $k \in \mathbb{N}$ and let $\Delta_k^2 = \Delta_k \times \Delta_k$ where Δ_k is the simplex $\{\mathbf{p} = (p_1, \ldots, p_k) : p_i \geq 0,\ \sum_1^k p_i = 1\}$. The notation for a generic point of Δ_k^2 will be $\mathbf{p} = (\mathbf{p}^{(1)}, \mathbf{p}^{(2)})$ where each $\mathbf{p}^{(i)} = (p_1^{(i)}, \ldots, p_k^{(i)}) \in \Delta_k$. We think of Δ_k^2 as a subset of $(\mathbb{R}^k)^2$, the

$2k$–dimensional space of all points $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$, where $\mathbf{x}^{(i)} = (x_1^{(i)}, \ldots, x_k^{(i)})$ with the $x_j^{(i)}$ real. The orthant in which all the $x_j^{(i)}$ are non-negative will be denoted $(\mathbb{R}_+^k)^2$. For $\mathbf{x}$, $\mathbf{y} \in (\mathbb{R}^k)^2$ we shall write $\mathbf{x}\mathbf{y}^\top$ for

$$\mathbf{x}^{(1)}\mathbf{y}^{(1)\top} + \mathbf{x}^{(2)}\mathbf{y}^{(2)\top} = \sum_{i=1}^{2}\sum_{j=1}^{k} x_j^{(i)} y_j^{(i)}.$$

Also we need a norm on $(\mathbb{R}^k)^2$: $\|\mathbf{x}\| = \|(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})\| := \|\mathbf{x}^{(1)}\|_1 \vee \|\mathbf{x}^{(2)}\|_1$, where

$$\|\mathbf{x}^{(i)}\|_1 := \sum_{j=1}^{k} |x_j^{(i)}| \qquad (i = 1, 2).$$

We cut the sides of A each into $k \geq 2$ equal-sized pieces, bounded by points

$$t_j^{(i)} := a^{(i)} + \frac{j}{k}(b^{(i)} - a^{(i)}) \qquad (j = 0, 1, \ldots, k;\ i = 1, 2),$$

and define a 'discretization' map $\mathbf{d}_k : D_R \to \Delta_k^2$ on the parametrized paths by $\mathbf{d}_k(\mathbf{f}) := (\mathbf{d}_k^{(1)}(f^{(1)}), \mathbf{d}_k^{(2)}(f^{(2)}))$, where $\mathbf{d}_k^{(i)}(f^{(i)}) := (d_{k1}^{(i)}, \ldots, d_{kk}^{(i)})$ and

$$\begin{aligned} d_{k1}^{(i)} &:= f^{(i)}(t_1^{(i)}), \\ d_{kj}^{(i)} &:= f^{(i)}(t_j^{(i)}) - f^{(i)}(t_{j-1}^{(i)}) \qquad (j = 2, \ldots, k), \end{aligned}$$

for $i = 1, 2$.

The result of applying the discretization map to the sample parametrized path $\mathbf{M}_n$ is that from the 1–component empirical d.f. $M_n^{(1)}$ we get the proportions of the 1–component of the observations in the cells $[a^{(1)}, t_1^{(1)}]$, $(t_1^{(1)}, t_2^{(1)}]$, $\ldots$, $(t_{k-1}^{(1)}, b^{(1)}]$, and similarly for the 2–component. The image $\mathbf{d}_k(\mathbf{M}_n)$ lies in the subset Δ_{kn}^2 of Δ_k^2 consisting of all $\mathbf{p} = (\mathbf{p}^{(1)}, \mathbf{p}^{(2)}) \in \Delta_k^2$ such that the entries $p_j^{(i)}$ of $\mathbf{p}^{(i)} = (p_1^{(i)}, \ldots, p_k^{(i)})$ are rationals with denominator n. In Lemma 3.6 below we will estimate the probability that $\mathbf{d}_k(\mathbf{M}_n)$ equals a generic element of this subset.

Now the above subdivision of the sides of A divides A into k^2 congruent cells

$$A_{uv} := (t_{u-1}^{(1)}, t_u^{(1)}] \times (t_{v-1}^{(2)}, t_v^{(2)}] \qquad (u, v = 1, 2, \ldots, k).$$

Define $\overline{h}_k$, $\underline{h}_k : A \to \mathbb{R}$ by

$$\begin{aligned} \overline{h}_k(\mathbf{z}) &:= \sum_{u=1}^{k}\sum_{v=1}^{k} \mathbf{1}\{\mathbf{z} \in A_{uv}\} h^\vee(A_{uv}) \qquad (\mathbf{z} \in A), \\ \underline{h}_k(\mathbf{z}) &:= \sum_{u=1}^{k}\sum_{v=1}^{k} \mathbf{1}\{\mathbf{z} \in A_{uv}\} h^\wedge(A_{uv}) \qquad (\mathbf{z} \in A). \end{aligned}$$

For any $\mathbf{p} \in \Delta_k^2$, let $s_m^{(i)} = s_m^{(i)}(\mathbf{p}) := \sum_{l=1}^m p_l^{(i)}$ for $i = 1, 2$ and $m = 0, 1, \dots, k$, and define functions $\overline{S}_k$, $\underline{S}_k$, U_k, $\overline{T}_k$, $\underline{T}_k : (\mathbb{R}^k)^2 \to [-\infty, \infty]$ as follows:

$$\overline{S}_k(\mathbf{p}) := \begin{cases} \sum_{u=1}^k \sum_{v=1}^k ((s_u^{(1)} \wedge s_v^{(2)}) - (s_{u-1}^{(1)} \vee s_{v-1}^{(2)}))^+ \log \frac{h^\vee(A_{uv})}{H(A)} & (\mathbf{p} \in \Delta_k^2), \\ -\infty & (\text{elsewhere}), \end{cases}$$

$$\underline{S}_k(\mathbf{p}) := \begin{cases} \sum_{u=1}^k \sum_{v=1}^k ((s_u^{(1)} \wedge s_v^{(2)}) - (s_{u-1}^{(1)} \vee s_{v-1}^{(2)}))^+ \log \frac{h^\wedge(A_{uv})}{H(A)} & (\mathbf{p} \in \Delta_k^2), \\ -\infty & (\text{elsewhere}), \end{cases}$$

$$U_k(\mathbf{p}) := \begin{cases} \sum_{i=1}^2 \sum_{j=1}^k p_j^{(i)} \log p_j^{(i)} + 2\log k & (\mathbf{p} \in (\mathbb{R}_+^k)^2), \\ \infty & (\text{elsewhere}), \end{cases}$$

$$\overline{T}_k := \overline{S}_k - U_k,$$

$$\underline{T}_k := \underline{S}_k - U_k.$$

When $p = 0$, interpret $p \log p$ as 0.

Lemma 3.5. *With inverses defined as in (2.3),*

$$\overline{S}_k(\mathbf{d}_k(\mathbf{f})) = \int_0^1 \log \frac{\overline{h}_k(\mathbf{f}^\leftarrow(p))}{H(A)}\, dp \qquad (\mathbf{f} \in D_R), \tag{3.3}$$

$$\underline{S}_k(\mathbf{d}_k(\mathbf{f})) = \int_0^1 \log \frac{\underline{h}_k(\mathbf{f}^\leftarrow(p))}{H(A)}\, dp \qquad (\mathbf{f} \in D_R). \tag{3.4}$$

For the special case when $\mathbf{f} = (f^{(1)}, f^{(2)})$ *is given by*

$$f^{(i)}(t) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{z_j^{(i)} \le t\} \qquad (i = 1, 2;\ t \in \mathbb{R}),$$

for some points $\mathbf{z}_1 < \mathbf{z}_2 < \cdots < \mathbf{z}_n$ *in* A*, we have*

$$\prod_{j=1}^n \frac{\overline{h}_k(\mathbf{z}_j)}{H(A)} = e^{n\overline{S}_k(\mathbf{d}_k(\mathbf{f}))}, \qquad \prod_{j=1}^n \frac{\underline{h}_k(\mathbf{z}_j)}{H(A)} = e^{n\underline{S}_k(\mathbf{d}_k(\mathbf{f}))}. \tag{3.5}$$

Proof: Fix $\mathbf{f} \in D_R$. As p increases through the interval $(0, 1]$ the values $\mathbf{f}^\leftarrow(p)$ enter, in order, a sequence $A_1, \dots, A_m$, say, of the sets A_{uv}. Thus there exist $0 = p_0 < p_1 < \cdots < p_m = 1$ such that $\mathbf{f}^\leftarrow(p) \in A_l$ for $p_{l-1} < p \le p_l$, for $l = 1, \dots, m$. Now those cells A_{uv} that are one of the sequence $A_1, \dots, A_m$ are characterized by there being an l such that $p_{l-1} = f^{(1)}(t_{u-1}^{(1)}) \vee f^{(2)}(t_{v-1}^{(2)})$ and $p_l = f^{(1)}(t_u^{(1)}) \wedge f^{(2)}(t_v^{(2)})$. The other cells are those not entered by the path and lying above it, for which $f^{(2)}(t_{v-1}^{(2)}) > f^{(1)}(t_u^{(1)})$, and those not entered by the path and lying below it, for which $f^{(1)}(t_{u-1}^{(1)}) > f^{(2)}(t_v^{(2)})$; in either case, $f^{(1)}(t_u^{(1)}) \wedge f^{(2)}(t_v^{(2)}) < f^{(1)}(t_{u-1}^{(1)}) \vee f^{(2)}(t_{v-1}^{(2)})$. So

$$\begin{aligned} &\overline{S}_k(\mathbf{d}_k(\mathbf{f})) \\ &\quad = \sum_{u=1}^k \sum_{v=1}^k \big((f^{(1)}(t_u^{(1)}) \wedge f^{(2)}(t_v^{(2)})) - (f^{(1)}(t_{u-1}^{(1)}) \vee f^{(2)}(t_{v-1}^{(2)}))\big)^+ \log \frac{h^\vee(A_{uv})}{H(A)} \\ &\quad = \sum_{l=1}^m (p_l - p_{l-1}) \log \frac{h^\vee(A_l)}{H(A)}, \end{aligned}$$

and (3.3) follows. (3.4) is similar.

For the first formula in (3.5) (the other is proved similarly) we have $\mathbf{f}^{\leftarrow}(p) = \sum_{j=1}^{n} \mathbf{z}_j \mathbf{1}\{(j-1)/n < p \le j/n\}$, so that by (3.3),

$$\overline{S}_k(\mathbf{d}_k(\mathbf{f})) = \sum_{j=1}^{n} \frac{1}{n} \log \frac{\overline{h}_k(\mathbf{z}_j)}{H(A)},$$

whence the formula. □

LEMMA 3.6. *Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be iid random vectors with common density given by (2.4). Then for every $\mathbf{p} \in \Delta^2_{kn}$,*

$$\left(\frac{\lambda(A)}{k^2}\right)^n \frac{e^{n\underline{S}_k(\mathbf{p})}}{\prod_{i=1}^2 \prod_{j=1}^k (np_j^{(i)})!} \le P(\mathbf{d}_k(\mathbf{M}_n) = \mathbf{p},\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n)$$
$$\le \left(\frac{\lambda(A)}{k^2}\right)^n \frac{e^{n\overline{S}_k(\mathbf{p})}}{\prod_{i=1}^2 \prod_{j=1}^k (np_j^{(i)})!}.$$

PROOF: We may write the event we are to bound the probability of as $\{(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in E_{n,\mathbf{p}}\}$, where

$$E_{n,\mathbf{p}} := \bigcap_{i=1}^{2} \bigcap_{j=1}^{k} \{(\mathbf{z}_1, \ldots, \mathbf{z}_n) \in (\mathbb{R}^2)^n : t^{(i)}_{j-1} < z^{(i)}_{ns^{(i)}_{j-1}+1} < z^{(i)}_{ns^{(i)}_{j-1}+2} < \cdots < z^{(i)}_{ns^{(i)}_{j}} \le t^{(i)}_j\}.$$

The said event thus has probability $\int_{E_{n,\mathbf{p}}} \prod_1^n (h(\mathbf{z}_i)\, d\mathbf{z}_i / H(A))$. By the previous lemma,

$$e^{n\underline{S}_k(\mathbf{p})} I_n \le P(\mathbf{d}_k(\mathbf{M}_n) = \mathbf{p},\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \le e^{n\overline{S}_k(\mathbf{p})} I_n,$$

where $I_n := \int_{E_{n,\mathbf{p}}} d\mathbf{z}_1 \ldots d\mathbf{z}_n$. And this last integral is easy to evaluate:

$$I_n = \prod_{i=1}^{2} \prod_{j=1}^{k} \int \cdots \int \mathbf{1}\{x_1, \ldots, x_{np^{(i)}_j} \in (t^{(i)}_{j-1}, t^{(i)}_j]\} \mathbf{1}\{x_1 < \cdots < x_{np^{(i)}_j}\}\, dx_1 \ldots dx_{np^{(i)}_j}$$
$$= \prod_{i=1}^{2} \prod_{j=1}^{k} \frac{(t^{(i)}_j - t^{(i)}_{j-1})^{np^{(i)}_j}}{(np^{(i)}_j)!} = \frac{(\lambda(A)/k^2)^n}{\prod_{i=1}^2 \prod_{j=1}^k (np^{(i)}_j)!}.$$

The conclusion follows. □

LEMMA 3.7. *Set*

$$h_{uv} := \log \frac{h^{\vee}(A_{uv})}{H(A)} \qquad (u, v = 1, \ldots, k),$$
$$\gamma_{uv} := h_{u+1,v+1} - h_{u+1,v} - h_{u,v+1} + h_{uv} \qquad (u, v = 1, \ldots, k).$$

Then for $\mathbf{p} \in \Delta_k^2$,

$$\overline{S}_k(\mathbf{p}) = R_k(\mathbf{p}) - \sum_{u=1}^{k-1}\sum_{v=1}^{k-1} \mathbf{1}\{s_v^{(2)} < s_u^{(1)}\}(s_u^{(1)} - s_v^{(2)})\gamma_{uv},$$

where

$$\begin{aligned} R_k(\mathbf{p}) &:= \sum_{u=1}^{k-1} s_u^{(1)}(h_{u1} - h_{u+1,1}) + \sum_{v=1}^{k-1} s_v^{(2)}(h_{kv} - h_{k,v+1}) + h_{kk} \\ &= \sum_{u=1}^{k-1} p_u^{(1)}(h_{11} - h_{u+1,1}) + \sum_{v=1}^{k-1} p_v^{(2)}(h_{k1} - h_{k,v+1}) + h_{kk}. \end{aligned}$$

Proof: We have

$$\begin{aligned} \overline{S}_k(\mathbf{p}) &= \sum_{u=1}^{k}\sum_{v=1}^{k} \mathbf{1}\{s_{v-1}^{(2)} < s_u^{(1)}\}\mathbf{1}\{s_{u-1}^{(1)} < s_v^{(2)}\}\big((s_u^{(1)} \wedge s_v^{(2)}) - (s_{u-1}^{(1)} \vee s_{v-1}^{(2)})\big)h_{uv} \\ &= \sum_{u=1}^{k}\sum_{v=1}^{k}\big(\mathbf{1}\{s_{u-1}^{(1)} < s_v^{(2)} < s_u^{(1)}\}s_v^{(2)} + \mathbf{1}\{s_{v-1}^{(2)} < s_u^{(1)} \le s_v^{(2)}\}s_u^{(1)} \\ &\quad - \mathbf{1}\{s_{v-1}^{(2)} < s_{u-1}^{(1)} < s_v^{(2)}\}s_{u-1}^{(1)} - \mathbf{1}\{s_{u-1}^{(1)} \le s_{v-1}^{(2)} < s_u^{(1)}\}s_{v-1}^{(2)}\big)h_{uv} \\ &= \sum_{u=1}^{k} s_u^{(1)} \sum_{v=1}^{k} \mathbf{1}\{s_{v-1}^{(2)} < s_u^{(1)} \le s_v^{(2)}\}h_{uv} - \sum_{u=1}^{k-1} s_u^{(1)} \sum_{v=1}^{k} \mathbf{1}\{s_{v-1}^{(2)} < s_u^{(1)} < s_v^{(2)}\}h_{u+1,v} \\ &\quad + \sum_{v=1}^{k-1} s_v^{(2)} \sum_{u=1}^{k} \mathbf{1}\{s_{u-1}^{(1)} < s_v^{(2)} < s_u^{(1)}\}h_{uv} - \sum_{v=1}^{k-1} s_v^{(2)} \sum_{u=1}^{k} \mathbf{1}\{s_{u-1}^{(1)} \le s_v^{(2)} < s_u^{(1)}\}h_{u,v+1} \\ &= \sum_{u=1}^{k-1} s_u^{(1)} \sum_{v=1}^{k} \mathbf{1}\{s_{v-1}^{(2)} < s_u^{(1)} \le s_v^{(2)}\}(h_{uv} - h_{u+1,v}) \\ &\quad + \sum_{v=1}^{k-1} s_v^{(2)} \sum_{u=1}^{k-1} \mathbf{1}\{s_{u-1}^{(1)} \le s_v^{(2)} < s_u^{(1)}\}(h_{uv} - h_{u,v+1}) + h_{kk}. \end{aligned} \tag{3.6}$$

But the claimed expression for $\overline{S}_k(\mathbf{p})$ equals, by telescoping,

$$\begin{aligned} &R_k(\mathbf{p}) + \sum_{v=1}^{k-1} s_v^{(2)} \sum_{u=1}^{k-1} \mathbf{1}\{s_{u-1}^{(1)} \le s_v^{(2)} < s_u^{(1)}\}(h_{uv} - h_{u,v+1} + h_{k,v+1} - h_{kv}) \\ &\qquad - \sum_{u=1}^{k-1} s_u^{(1)} \sum_{v=1}^{k-1} \mathbf{1}\{s_v^{(2)} < s_u^{(1)} \le s_{v+1}^{(2)}\}(h_{u+1,v+1} - h_{u,v+1} + h_{u1} - h_{u+1,1}) \\ &= R_k(\mathbf{p}) + \sum_{v=1}^{k-1} s_v^{(2)} \sum_{u=1}^{k} - \sum_{u=1}^{k-1} s_u^{(1)} \sum_{v=0}^{k-1} \end{aligned}$$

$$= R_k(\mathbf{p}) + \sum_{v=1}^{k-1} s_v^{(2)} \sum_{u=1}^{k} \mathbf{1}\{s_{u-1}^{(1)} \le s_v^{(2)} < s_u^{(1)}\}(h_{uv} - h_{u,v+1}) + \sum_{v=1}^{k-1} s_v^{(2)}(h_{k,v+1} - h_{kv})$$
$$- \sum_{u=1}^{k-1} s_u^{(1)} \sum_{v=1}^{k} \mathbf{1}\{s_{v-1}^{(2)} < s_u^{(1)} \le s_v^{(2)}\}(h_{u+1,v} - h_{uv}) - \sum_{u=1}^{k-1} s_u^{(1)}(h_{u1} - h_{u+1,1}),$$

which reduces, on inserting the defining formula for R_k, to the expression (3.6). $\square$

LEMMA 3.8. *Under the conditions of Theorem 3.2, all γ_{uv} are non-negative, and $\overline{S}_k$ is concave.*

PROOF: Without loss of generality we may assume that the normalization constant $H(A)$ in the defining formulae of the h_{uv} is 1. As $\log h$ is L-superadditive,

$$\log h(x^{(1)}, x^{(2)}) - \log h(x^{(1)}, y^{(2)}) - \log h(y^{(1)}, x^{(2)}) - \log h(y^{(1)}, y^{(2)}) \ge 0$$

for any $\mathbf{x} < \mathbf{y}$ in A.

Fix u, $v \in \{1, \dots, k\}$. Take the components $x^{(1)}$ and $y^{(2)}$ of $\mathbf{x}$ and $\mathbf{y}$ so that $\log h$ attains its supremum in $A_{u,v+1}$ at $(x^{(1)}, y^{(2)})$, and take the components $y^{(1)}$ and $x^{(2)}$ so that $\log h$ attains its supremum in $A_{u+1,v}$ at $(y^{(1)}, x^{(2)})$. Thus

$$\log h(\mathbf{x}) - h_{u,v+1} - h_{u+1,v} + \log h(\mathbf{y}) \ge 0.$$

But these choices put $\mathbf{x}$ in A_{uv} and $\mathbf{y}$ in $A_{u+1,v+1}$, so $\log h(\mathbf{x}) \le h_{uv}$ and $\log h(\mathbf{y}) \le h_{u+1,v+1}$, and we conclude $\gamma_{uv} \ge 0$ as claimed.

As the γ_{uv} are non-negative, the result of Lemma 3.7 may be written

$$\overline{S}_k(\mathbf{p}) = R_k(\mathbf{p}) + \sum_{u=1}^{k-1} \sum_{v=1}^{k-1} \min((s_v^{(2)} - s_u^{(1)})\gamma_{uv}, 0).$$

Now R_k is an affine function, so concave. In the sum, each summand is the pointwise minimum of two linear functions of $\mathbf{p}$, so concave. So $\overline{S}_k$ is here expressed as a finite sum of concave functions, and is thus itself concave. $\square$

Remark 3.9. If any of the γ_{uv} are negative then we can prove $\overline{S}_k$ is not concave and our proof of Theorem 3.2 collapses. We doubt if there is a unique limit curve without concavity of $\overline{S}_k$ at some fine enough k. Essentially, this is the condition that $\log h$ be L-superadditive.

LEMMA 3.10. *Under the conditions of Theorem 3.2, $\overline{T}_k$ has a unique maximum at a point $\mathbf{q}_k = (\mathbf{q}_k^{(1)}, \mathbf{q}_k^{(2)})$ in $\mathrm{ri}(\Delta_k^2)$, the relative interior of Δ_k^2. On the components of $\mathbf{q}_k^{(i)} = (q_{k1}^{(i)}, \dots, q_{kk}^{(i)})$ we have the bounds*

$$\frac{m}{m + (k-m)e^{6C_h k/\max(m,k-m)}} \le \sum_{j \in B} q_{kj}^{(i)}$$
$$\le \frac{m}{m + (k-m)e^{-6C_h k/\max(m,k-m)}} \qquad (i = 1, 2), \qquad (3.7)$$

where B is any set of m elements of $\{1,\ldots,k\}$, and

$$C_h := \sup_{\mathbf{z}\in A}\left|\log\frac{h(\mathbf{z})}{H(A)}\right|.$$

PROOF: We need the following $(2k-2)$–dimensional subspace of $(\mathbb{R}^k)^2$:

$$V_k := \{\mathbf{x} = (\mathbf{x}^{(1)},\mathbf{x}^{(2)}) \in (\mathbb{R}^k)^2 : \mathbf{x}^{(1)}\mathbf{1}^\top = 0,\ \mathbf{x}^{(2)}\mathbf{1}^\top = 0\},$$

where $\mathbf{1}$ is the row k–tuple of 1s. Its 2–dimensional orthogonal complement is

$$W_k := \{\mathbf{x} = (u\mathbf{1}, v\mathbf{1}) : u, v \in \mathbb{R}\}.$$

Note that Δ_k^2 is a convex subset of the affine flat $k^{-1}(\mathbf{1},\mathbf{1}) + V_k$ in $(\mathbb{R}^k)^2$.

Pick u, $v \in \{1,\ldots,k\}$. The set of $\mathbf{p}\in\Delta_k^2$ for which $s_v^{(2)} - s_u^{(1)} = 0$ is the restriction of the hyperplane $\{\mathbf{x}\in(\mathbb{R}^k)^2 : \sum_{w=1}^u x_w^{(1)} - \sum_{w=1}^v x_w^{(2)} = 0\}$ in $(\mathbb{R}^k)^2$ to Δ_k^2. These hyperplanes for the various u, v decompose the polyhedral closed convex set Δ_k^2 into a finite union of polyhedral closed convex sets $\Delta_k^2(1)$, ..., $\Delta_k^2(m_k)$ say, with disjoint relative interiors. Pick $\mathbf{p}$ in one of these relative interiors, then $s_u^{(1)} \neq s_v^{(2)}$ for all u, $v = 1$, ..., k. For sufficiently small (in norm) $\boldsymbol{\delta}\in V_k$ the indicator functions in Lemma 3.7's expansions of $\overline{S}_k(\mathbf{p})$ and $\overline{S}_k(\mathbf{p}+\boldsymbol{\delta})$ have the same values as each other, so

$$\overline{S}_k(\mathbf{p}+\boldsymbol{\delta}) = R_k(\mathbf{p}+\boldsymbol{\delta}) - \sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\mathbf{1}\{s_v^{(2)} < s_u^{(1)}\}\Big(s_u^{(1)} + \sum_{w=1}^{u}\delta_w^{(1)} - s_v^{(2)} - \sum_{w=1}^{v}\delta_w^{(2)}\Big)\gamma_{uv}.$$

Thus

$$\begin{aligned}
&\overline{S}_k(\mathbf{p}+\boldsymbol{\delta}) - \overline{S}_k(\mathbf{p})\\
&\quad= R_k(\mathbf{p}+\boldsymbol{\delta}) - R_k(\mathbf{p}) - \sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\mathbf{1}\{s_v^{(2)} < s_u^{(1)}\}\Big(\sum_{w=1}^{u}\delta_w^{(1)} - \sum_{w=1}^{v}\delta_w^{(2)}\Big)\gamma_{uv}\\
&\quad= \sum_{w=1}^{k-1}(h_{11} - h_{w+1,1})\delta_w^{(1)} + \sum_{w=1}^{k-1}(h_{k1} - h_{k,w+1})\delta_w^{(2)}\\
&\qquad - \sum_{w=1}^{k-1}\sum_{u=w}^{k-1}\sum_{v=1}^{k-1}\mathbf{1}\{s_v^{(2)} < s_u^{(1)}\}\gamma_{uv}\delta_w^{(1)} + \sum_{w=1}^{k-1}\sum_{u=1}^{k-1}\sum_{v=w}^{k-1}\mathbf{1}\{s_v^{(2)} < s_u^{(1)}\}\gamma_{uv}\delta_w^{(2)}.
\end{aligned}$$

The terms γ_{uv} are non-negative and so any partial sum of them lies between zero and

$$\begin{aligned}
\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\gamma_{uv} &= \sum_{u=1}^{k-1}(h_{u1} - h_{uk} + h_{u+1,k} - h_{u+1,1})\\
&= h_{11} - h_{k1} - h_{1k} + h_{kk} \le 4C_h
\end{aligned}$$

Thus we may rewrite the above as

$$\overline{S}_k(\mathbf{p}+\boldsymbol{\delta}) - \overline{S}_k(\mathbf{p}) = \sum_{i=1}^{2}\sum_{j=1}^{k} \eta_j^{(i)\prime}\delta_j^{(i)},$$

where the $\eta_j^{(i)\prime}$ are bounded in absolute value by $6C_h$. Further, if we set $\eta_j^{(i)} := \eta_j^{(i)\prime} - k^{-1}\sum_{l=1}^{k}\eta_l^{(i)\prime}$ then the $\eta_j^{(i)}$ also have absolute values bounded by $6C_h$, and we have

$$\overline{S}_k(\mathbf{p}+\boldsymbol{\delta}) - \overline{S}_k(\mathbf{p}) = \sum_{i=1}^{2}\sum_{j=1}^{k} \eta_j^{(i)}\delta_j^{(i)},$$

since $\boldsymbol{\delta} \in V_k$. Thus

$$\overline{S}_k(\mathbf{p}+\boldsymbol{\delta}) - \overline{S}_k(\mathbf{p}) = \boldsymbol{\eta}(l)^\top\boldsymbol{\delta} \qquad (\mathbf{p}, \mathbf{p}+\boldsymbol{\delta} \in \Delta_k^2(l)),$$

where $\boldsymbol{\eta}(l) = (\boldsymbol{\eta}^{(1)}(l), \boldsymbol{\eta}^{(2)}(l)) \in V_k$ and has components bounded in absolute value by $6C_h$. In terms of the space $S_k^2 := \{\mathbf{x} \in (\mathbb{R}^k)^2 : \|x\| = 1\}$ this can alternatively be written

$$\overline{S}_k(\mathbf{p}+t\mathbf{d}) - \overline{S}_k(\mathbf{p}) = t\boldsymbol{\eta}(l)^\top\mathbf{d} \qquad (\mathbf{p} \in \Delta_k^2(l)) \tag{3.8}$$

for every $\mathbf{d} \in S_k^2 \cap V_k$, and every sufficiently small $t > 0$.

The remaining $\mathbf{p}$ in $\mathrm{ri}(\Delta_k^2)$ are those on the boundaries of several of the $\Delta_k^2(l)$, and for those $\mathbf{p}$ (3.8) still holds but with $\boldsymbol{\eta}(l)$ depending on $\mathbf{d}$, in that the value of l is any of those for which $\mathbf{p}+t\mathbf{d} \in \Delta_k^2(l)$. Thus at any point $\mathbf{p} \in \mathrm{ri}(\Delta_k^2)$, the derivative of $\overline{S}_k$ in any direction $\mathbf{d} \in S_k^2 \cap V_k$ is $\boldsymbol{\eta}(l)^\top\mathbf{d}$, where l is the index for which $\mathbf{p}+t\mathbf{d} \in \Delta_k^2(l)$ for small $t > 0$. Since $-\overline{S}_k$ is convex, and equal to ∞ off Δ_k^2, we conclude that its subdifferential (cf. [Rockafellar 1970, §23]) $\partial(-\overline{S}_k)(\mathbf{p})$ at $\mathbf{p}$ is the set of convex combinations of these directional derivatives and of all vectors in W_k. Thus

$$\partial(-\overline{S}_k)(\mathbf{p}) = M_k(\mathbf{p}) + W_k \qquad (\mathbf{p} \in \mathrm{ri}(\Delta_k^2)), \tag{3.9}$$

where $M_k(\mathbf{p})$ is a convex subset of V_k, all of whose elements $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)})$ have their components $\eta_j^{(i)}$ bounded in absolute value by $6C_h$.

The convex function $-\overline{T}_k$ on $(\mathbb{R}^k)^2$ has subdifferential

$$\partial(-\overline{T}_k)(\mathbf{p}) = \partial(-\overline{S}_k)(\mathbf{p}) + \partial U_k(\mathbf{p}) \qquad (\mathbf{p} \in (\mathbb{R}^k)^2)$$

by [Rockafellar 1970, Theorem 23.8]. For $\mathbf{p} \in \mathrm{ri}(\Delta_k^2)$ we have by [Rockafellar 1970, Theorem 25.1], and (3.9), that

$$\partial(-\overline{T}_k)(\mathbf{p}) = M_k(\mathbf{p}) + W_k + \{\nabla U_k(\mathbf{p})\}.$$

The j^{th} component of the i^{th} sub-vector of $\nabla U_k(\mathbf{p})$ is $1 + \log p_j^{(i)}$. Let $\mathbf{r}(\mathbf{p}) = (\mathbf{r}^{(1)}(\mathbf{p}), \mathbf{r}^{(2)}(\mathbf{p}))$ be the element of V_k with components

$$r_j^{(i)}(\mathbf{p}) := -\log p_j^{(i)} + k^{-1}\sum_{l=1}^{k}\log p_l^{(i)} \qquad (i = 1, 2;\ j = 1, \ldots, k),$$

then we can re-express the subdifferential in orthogonal-decomposition form

$$\partial(-\overline{T}_k)(\mathbf{p}) = M_k(\mathbf{p}) + \{-\mathbf{r}(\mathbf{p})\} + W_k \qquad (\mathbf{p} \in \mathrm{ri}(\Delta_k^2)), \tag{3.10}$$

where $M_k(\mathbf{p}) + \{-\mathbf{r}(\mathbf{p})\} \subset V_k$.

Since $\overline{S}_k$ is concave and U_k is strictly concave, $\overline{T}_k = \overline{S}_k - U_k$ has a unique maximum at $\mathbf{q}_k$, say, where $\mathbf{q}_k \in \Delta_k^2$. In fact $\mathbf{q}_k \in \mathrm{ri}(\Delta_k^2)$ as we now show. For the restriction of $\overline{T}_k$ to Δ_k^2 is continuous. Suppose we pick any point $\mathbf{p}_0$ on the relative boundary of Δ_k^2 and a straight-line path from $\mathbf{p}_0$ into $\mathrm{ri}(\Delta_k^2)$. Since $\mathbf{p}_0 = (\mathbf{p}_0^{(1)}, \mathbf{p}_0^{(2)})$ has one or more components $p_{0j}^{(i)}$ equal to 0, the corresponding logarithms in the expression for $r_j^{(i)}(\mathbf{p})$ will be large at points $\mathbf{p}$ on the path close to $\mathbf{p}_0$, and will dominate the contributions from the uniformly bounded sets $M_k(\mathbf{p})$ in the expressions for $\partial(-\overline{T}_k)(\mathbf{p})$ for the various $\mathbf{p}$. Hence it is straightforward to calculate that, close to $\mathbf{p}_0$ on the path, the direction of the path has negative scalar product with all members of the set $\partial(-\overline{T}_k)(\mathbf{p})$; that is, $-\overline{T}_k$ decreases as we move on the path from $\mathbf{p}_0$ into $\mathrm{ri}(\Delta_k^2)$. So $\mathbf{q}_k$ cannot be $\mathbf{p}_0$, and is thus in $\mathrm{ri}(\Delta_k^2)$.

From the definition of subdifferential we obtain that $\mathbf{0} \in \partial(-\overline{T}_k)(\mathbf{q}_k)$, and since $\mathbf{q}_k \in \mathrm{ri}(\Delta_k^2)$ we find from (3.10) and orthogonality that $\mathbf{0} \in M_k(\mathbf{q}_k) + \{-\mathbf{r}(\mathbf{q}_k)\}$, that is, $\mathbf{r}(\mathbf{q}_k) \in M_k(\mathbf{q}_k)$. So the absolute values of the components of $\mathbf{r}(\mathbf{q}_k)$ are bounded by $6C_h$. (3.7) then follows from Lemma 3.11 below, with $c := 6C_h$. $\square$

Lemma 3.11. *Fix $c > 0$ and integers $k \geq m \geq 1$. Let $p_1, \ldots, p_k$ be positive reals satisfying*

$$\left| \log p_j - \frac{1}{k} \sum_{l=1}^{k} \log p_l \right| \leq c \qquad (j = 1, \ldots, k) \tag{3.11}$$

and

$$\sum_{l=1}^{k} p_l = 1. \tag{3.12}$$

Then

$$\frac{m}{m + (k-m)e^{ck/\max(m,k-m)}} \leq \sum_{j=1}^{m} p_j \leq \frac{m}{m + (k-m)e^{-ck/\max(m,k-m)}}. \tag{3.13}$$

Remark 3.12. The bounds are sharp, in that by taking $p_1 = \cdots = p_m := a$, and $p_{m+1} = \cdots = p_k := b$, and choosing a and b appropriately, either bound can be satisfied, within the constraints, whatever the values of k and m.

Proof: First observe that the constraints bound all the p_j away from 0. For instance, (3.11) gives that $\log p_j \leq k^{-1} \log p_1 + c$ for $j = 2, \ldots, k$, hence

$$\begin{aligned} 1 = \sum_{1}^{k} p_j &\leq p_1 + (k-1)e^c p_1^{1/k} \\ &\leq (1 + (k-1)e^c) p_1^{1/k}, \end{aligned}$$

which is a lower bound on p_1.

Consider the following problem in the variables $p_1, \dots, p_k$: minimize $s := \sum_1^m p_j$ subject to

$$\log p_j - \frac{1}{k}\sum_{l=1}^{k} \log p_l \le c \qquad (j = 1, \dots, k), \tag{3.14}$$

and (3.12). Standard constrained-optimization techniques reveal that the constraints (3.14) for $j = m+1, \dots, k$ must be non-slack at the minimum, so the variables $p_{m+1}, \dots, p_k$ are equal, necessarily to $(1-s)/(k-m)$. We then find that each inequality (3.14), for $j = m+1, \dots, k$, becomes

$$\log \frac{1-s}{k-m} - \frac{1}{k}\sum_{l=1}^{m} \log p_l - \frac{k-m}{k} \log \frac{1-s}{k-m} \le c,$$

that is,

$$\frac{1}{m}\sum_{l=1}^{m} \log p_l \ge \log \frac{1-s}{k-m} - c\frac{k}{m}.$$

Exponentiate and apply the Arithmetic Mean-Geometric Mean inequality: thus

$$\frac{s}{m} \ge \frac{1-s}{k-m} e^{-ck/m}.$$

Rearranging this gives

$$s \ge \frac{m}{m + (k-m)e^{ck/m}}. \tag{3.15}$$

In the similar constrained-optimization problem of minimizing $s := \sum_1^m p_j$ subject to

$$\log p_j - \frac{1}{k}\sum_{l=1}^{k} \log p_l \ge -c \qquad (j = 1, \dots, k), \tag{3.16}$$

and (3.12), it again is straightforward to show that, at the minimum, all of $p_{m+1}, \dots, p_k$ must equal $(1-s)/(k-m)$. The first m inequalities of (3.11) then give

$$k \log p_j - \sum_{l=1}^{m} \log p_l - (k-m) \log \frac{1-s}{k-m} \ge -ck \qquad (j = 1, \dots, m).$$

Add these, obtaining

$$\frac{1}{m}\sum_{l=1}^{m} \log p_l \ge \log \frac{1-s}{k-m} - \frac{ck}{k-m}.$$

The same calculations as above then give

$$s \ge \frac{m}{m + (k-m)e^{ck/(k-m)}}.$$

This and (3.15) give the left-hand inequality of (3.13).

With (3.12), the left-hand inequality of (3.13) gives

$$\sum_{j=m+1}^{k} p_j \le 1 - \frac{m}{m + (k-m)e^{ck/\max(m,k-m)}} = \frac{k-m}{k - m + me^{-ck/\max(m,k-m)}}.$$

Replace m by $k-m$ to get the right-hand inequality of (3.13). $\square$

LEMMA 3.13. *For any* $\mathbf{p} \in \Delta_k^2$,

$$\overline{T}_k(\mathbf{p}) - \overline{T}_k(\mathbf{q}_k) \ \leq -\tfrac{1}{4}\|\mathbf{p} - \mathbf{q}_k\|^2,$$

where $\mathbf{q}_k$ *is the location of the unique maximum of* $\overline{T}_k$, *as in Lemma 3.10.*

PROOF: As noted in the proof of Lemma 3.10,

$$\mathbf{0} \in \partial(-\overline{T}_k)(\mathbf{q}_k) = \partial(-\overline{S}_k)(\mathbf{q}_k) + \{\nabla U_k(\mathbf{q}_k)\}.$$

So $\nabla U_k(\mathbf{q}_k) \in \partial \overline{S}_k(\mathbf{q}_k)$, that is, $\overline{S}_k(\mathbf{p}) \leq \overline{S}_k(\mathbf{q}_k) + (\nabla U_k(\mathbf{q}_k))^\top(\mathbf{p} - \mathbf{q}_k)$. Re-arranging, and adding $U_k(\mathbf{q}_k) - U_k(\mathbf{p})$ to both sides, we obtain

$$\overline{T}_k(\mathbf{p}) - \overline{T}_k(\mathbf{q}_k) \leq (\nabla U_k(\mathbf{q}_k))^\top(\mathbf{p} - \mathbf{q}_k) - U_k(\mathbf{p}) + U_k(\mathbf{q}_k).$$

Let $HU_k(\mathbf{p})$ be the Hessian matrix (matrix of 2^{nd}-order partial derivatives) of U_k at $\mathbf{p}$. Then, by a two-term Taylor expansion of U_k about $\mathbf{q}_k$,

$$\overline{T}_k(\mathbf{p}) - \overline{T}_k(\mathbf{q}_k) \leq -\tfrac{1}{2}(\mathbf{p} - \mathbf{q}_k)^\top HU_k(\mathbf{p}')(\mathbf{p} - \mathbf{q}_k),$$

where $\mathbf{p}'$ is a point on the line segment between $\mathbf{q}_k$ and $\mathbf{p}$. Since $HU_k(\mathbf{p}) = \operatorname{diag}(1/p_j^{(i)})$ we conclude that

$$\overline{T}_k(\mathbf{p}) - \overline{T}_k(\mathbf{q}_k) \leq -\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{k}\frac{(p_j^{(i)} - q_{kj}^{(i)})^2}{p_j^{(i)\prime}},$$

where $\mathbf{p}' = (\mathbf{p}^{(1)\prime}, \mathbf{p}^{(2)\prime})$ has components $p_j^{(i)\prime}$. Finally,

$$\begin{aligned}\sum_{i=1}^{2}\sum_{j=1}^{k}\frac{(p_j^{(i)} - q_{kj}^{(i)})^2}{p_j^{(i)\prime}} &\geq \sum_{i=1}^{2}\sum_{j=1}^{k}(p_j^{(i)} - q_{kj}^{(i)})^2 \\ &\geq \frac{1}{2}\left(\sum_{i=1}^{2}\sum_{j=1}^{k}|p_j^{(i)} - q_{kj}^{(i)}|\right)^2,\end{aligned}$$

by the Cauchy-Schwarz inequality. The result follows. □

LEMMA 3.14. *Let* $|\mathbf{p}| := \max_{i=1,2}\max_{j=1,\ldots,k} p_j^{(i)}$ *for* $\mathbf{p} \in \Delta_k^2$. *Then for* $\mathbf{f}, \mathbf{g} \in D_R$,

$$\rho_L(\mathbf{f}, \mathbf{g}) \leq \rho_u(\mathbf{f}, \mathbf{g}) \leq \|\mathbf{d}_k(\mathbf{f}) - \mathbf{d}_k(\mathbf{g})\| + |\mathbf{d}_k(\mathbf{f})| + |\mathbf{d}_k(\mathbf{g})|.$$

PROOF: This is straightforward, from the definitions, and is left to the reader. □

The map $\mathbf{d}_k$ has no unique inverse but it will be convenient to define $\mathbf{c}_k : \Delta_k^2 \to D_R$ as follows. For $\mathbf{p} = (\mathbf{p}^{(1)}, \mathbf{p}^{(2)})$, where $\mathbf{p}^{(i)} = (p_1^{(i)}, \ldots, p_k^{(i)})$, set

$$\mathbf{c}_k(\mathbf{p}) := (c_k^{(1)}(\mathbf{p}^{(1)}), c_k^{(2)}(\mathbf{p}^{(2)})),$$

with

$$c_k^{(i)}(\mathbf{p}^{(i)})(t) := \begin{cases} 0 & (t \le a^{(i)}), \\ \sum_{l=1}^{j-1} p_l^{(i)} + p_j^{(i)} \frac{t - t_{j-1}^{(i)}}{t_j^{(i)} - t_{j-1}^{(i)}} & (t_{j-1}^{(i)} \le t \le t_j^{(i)};\ j = 1, \dots, k), \\ 1 & (t \ge b^{(i)}), \end{cases}$$

for $i = 1, 2$, so that $c_k^{(i)}(\mathbf{p}^{(i)})(\cdot)$ is the function with value $\sum_{l=1}^{j} p_l^{(i)}$ at $t_j^{(i)}$, for $j = 0$, $1, \dots, k$, and linearly interpolated in between. We then have

$$\mathbf{d}_k(\mathbf{c}_k(\mathbf{p})) = \mathbf{p} \qquad (\mathbf{p} \in \Delta_k^2). \tag{3.17}$$

On the other hand $\mathbf{c}_k(\mathbf{d}_k(\mathbf{f}))$ is a smoothed version of $\mathbf{f}$.

Proof of Theorem 3.2: From the inequality form of Stirling's formula (cf. e.g. [Feller 1970, II (9.15)]) we have

$$n \log n - n < \log n! < 1 + \tfrac{1}{2}\log(2\pi) + (n + \tfrac{1}{2}) \log n - n$$

for positive integers n, and indeed for non-integers if we interpret $n!$ as $\Gamma(n+1)$. Replace n by np where $p \in [0, 1]$, hence

$$np \log n + np \log p - np \le \log(np)! < 1 + \tfrac{1}{2}\log(2\pi) + (np + \tfrac{1}{2}) \log n + np \log p - np.$$

From Lemma 3.6 we thus have that for $\mathbf{p} \in \Delta_{kn}^2$,

$$\begin{aligned} &-2n \log n + 2n + n \log \lambda(A) + n\underline{T}_k(\mathbf{p}) - k \log n - k(2 + \log(2\pi)) \\ &\qquad \le \log P(\mathbf{d}_k(\mathbf{M}_n) = \mathbf{p},\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ &\qquad \le -2n \log n + 2n + n \log \lambda(A) + n\overline{T}_k(\mathbf{p}). \end{aligned} \tag{3.18}$$

Choose $\varepsilon > 0$ and set

$$\delta := \frac{1}{8}\left(\frac{\varepsilon}{8}\right)^2. \tag{3.19}$$

Choose an integer k so large that

$$\log h^{\vee}(A_{uv}) - \log h^{\wedge}(A_{uv}) < \tfrac{1}{2}\delta \qquad (u, v = 1, \dots, k) \tag{3.20}$$

and

$$\frac{1}{1 + (k-1)e^{-6C_h k/(k-1)}} < \frac{\varepsilon}{8}. \tag{3.21}$$

From (3.20), (3.3) and (3.4) it follows that

$$\overline{S}_k(\mathbf{p}) - \underline{S}_k(\mathbf{p}) < \tfrac{1}{2}\delta \qquad (\mathbf{p} \in \Delta_k^2). \tag{3.22}$$

With k chosen, we have from Lemma 3.10 that $\overline{T}_k$ has a unique maximum at some $\mathbf{q}_k \in \mathrm{ri}(\Delta_k^2)$. Since the restriction of $\overline{T}_k$ to Δ_k^2 is continuous at $\mathbf{q}_k$ we may choose $\varepsilon' > 0$ so small that

$$\overline{T}_k(\mathbf{p}) > \overline{T}_k(\mathbf{q}_k) - \tfrac{1}{2}\delta \tag{3.23}$$

for all $\mathbf{p}$ in the ball $B_{\varepsilon'}(\mathbf{q}_k) := \{\mathbf{p} \in \Delta_k^2 : \|\mathbf{p} - \mathbf{q}_k\| \le \varepsilon'\}$. We also make $\varepsilon' < \varepsilon/8$ and small enough that $B_{\varepsilon'}(\mathbf{q}_k) \subset \mathrm{ri}(\Delta_k^2)$. By (3.22) and (3.23), $\underline{T}_k(\mathbf{p}) > \overline{T}_k(\mathbf{q}_k) - \delta$ for all such $\mathbf{p}$. So from (3.18),

$$\begin{aligned} P(\|\mathbf{d}_k(\mathbf{M}_n) - \mathbf{q}_k\| \le \varepsilon',\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ > \sum_{\mathbf{p} \in \Delta_{kn}^2 \cap B_{\varepsilon'}(\mathbf{q}_k)} \exp\bigl(-2n\log n + 2n + n\log\lambda(A) + \\ + n(\overline{T}_k(\mathbf{q}_k) - \delta) - k\log n - k(2 + \log(2\pi))\bigr) \\ > (n\varepsilon')^{2k} \exp\bigl(-2n\log n + 2n + n\log\lambda(A) + \\ + n(\overline{T}_k(\mathbf{q}_k) - \delta) - k\log n - k(2 + \log(2\pi))\bigr), \end{aligned}$$

since there are at least $(n\varepsilon')^{2k}$ points of Δ_{kn}^2 within $B_{\varepsilon'}(\mathbf{q}_k)$.

On the other hand, from the other half of (3.18),

$$\begin{aligned} P(\|\mathbf{d}_k(\mathbf{M}_n) - \mathbf{q}_k\| > \tfrac{1}{8}\varepsilon,\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ < \sum_{\mathbf{p} \in \Delta_{kn}^2 \setminus B_{\varepsilon/8}(\mathbf{q}_k)} \exp\left(-2n\log n + 2n + n\log\lambda(A) + n\overline{T}_k(\mathbf{p})\right) \\ < \sum_{\mathbf{p} \in \Delta_{kn}^2 \setminus B_{\varepsilon/8}(\mathbf{q}_k)} \exp\left(-2n\log n + 2n + n\log\lambda(A) + n(\overline{T}_k(\mathbf{q}_k) - 2\delta)\right) \end{aligned}$$

(by Lemma 3.13 and (3.19))

$$\begin{aligned} &< n^{2k} \exp\left(-2n\log n + 2n + n\log\lambda(A) + n\overline{T}_k(\mathbf{q}_k) - 2n\delta\right) \\ &= \mathrm{o}(P(\|\mathbf{d}_k(\mathbf{M}_n) - \mathbf{q}_k\| \le \varepsilon',\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n)) \qquad (n \to \infty). \end{aligned}$$

It follows that

$$P(\|\mathbf{d}_k(\mathbf{M}_n) - \mathbf{q}_k\| \le \tfrac{1}{8}\varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \to 1 \qquad (n \to \infty).$$

Write $\mathbf{m}_\varepsilon := \mathbf{c}_k(\mathbf{q}_k)$. Then $\mathbf{d}_k(\mathbf{m}_\varepsilon) = \mathbf{q}_k$ by (3.17). By Lemma 3.10 and (3.21) we know $q_{kj}^{(i)} < \varepsilon/8$ for all $i = 1,\ 2$ and $j = 1, \ldots, k$; that is, $|\mathbf{d}_k(\mathbf{m}_\varepsilon)| < \varepsilon/8$ in the notation of Lemma 3.14. If $\|\mathbf{d}_k(\mathbf{M}_n) - \mathbf{d}_k(\mathbf{m}_\varepsilon)\| \le \varepsilon/8$ then it follows that $|\mathbf{d}_k(\mathbf{M}_n)| < \varepsilon/4$ and so, by Lemma 3.14, $\rho_L(\mathbf{M}_n, \mathbf{m}_\varepsilon) \le \varepsilon/8 + \varepsilon/8 + \varepsilon/4 = \varepsilon/2$. Thus

$$\begin{aligned} P(\rho_L(\mathbf{M}_n, \mathbf{m}_\varepsilon) \le \tfrac{1}{2}\varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ \ge P(\|\mathbf{d}_k(\mathbf{M}_n) - \mathbf{d}_k(\mathbf{m}_\varepsilon)\| \le \tfrac{1}{8}\varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ \to 1 \qquad (n \to \infty). \end{aligned} \tag{3.24}$$

Since (D_R, ρ_L) is a compact space, the sequence $(\mathbf{m}_{1/j})_{j=1,2,\ldots}$ has a subsequence $\mathbf{m}_{1/j'}$ convergent to $\mathbf{m}$, say. Then if we pick j_0' so large that $\rho_L(\mathbf{m}_{1/j_0'}, \mathbf{m}) \le \varepsilon/2$ we conclude finally that

$$\begin{aligned} P(\rho_L(\mathbf{M}_n, \mathbf{m}) \le \varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \ge P(\rho_L(\mathbf{M}_n, \mathbf{m}_{1/j_0'}) \le \tfrac{1}{2}\varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ \to 1 \qquad (n \to \infty), \end{aligned}$$

as desired. $\square$

4. Characterization of the limit.

For probability measures μ, ν on a common measurable space the *I-divergence* (Kullback-Leibler information number, relative entropy) is

$$D(\mu\|\nu) := \begin{cases} \int d\mu \log \frac{d\mu}{d\nu}, & \text{if } \mu \ll \nu, \\ \infty, & \text{if not.} \end{cases}$$

(cf. [Csiszár 1975]). Note that $D(\mu\|\nu) \geq 0$.

THEOREM 4.1. *Assume the conditions of Theorem 3.2, and define $T : D_R \to [-\infty, \infty)$ by*

$$T(\mathbf{f}) := \int_0^1 \log \frac{h(\mathbf{f}^{\leftarrow}(p))}{H(A)}\, dp - D(\mu(\mathbf{f})\|\lambda_A),$$

where λ_A is the uniform probability law on A. Then in Theorem 3.2 $\mathbf{m}$ is the unique element of D_R that maximizes T. Furthermore, we have the following.

(a) *The convergence can be strengthened to uniform: for each $\varepsilon > 0$,*

$$P(\rho_u(\mathbf{M}_n, \mathbf{m}) > \varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \to 0 \qquad (n \to \infty). \tag{4.1}$$

(b) *$\mathbf{m}$ is continuous, and $m^{(i)}(a^{(i)}) = 0$ and $m^{(i)}(b^{(i)}) = 1$, for $i = 1, 2$.*

(c) *For $i = 1, 2$, and any $a^{(i)} \leq t \leq t' \leq b^{(i)}$,*

$$\begin{aligned} \frac{\delta}{\delta + (1-\delta)e^{6C_h/\max(\delta, 1-\delta)}} &\leq m^{(i)}(t') - m^{(i)}(t) \\ &\leq \frac{\delta}{\delta + (1-\delta)e^{-6C_h/\max(\delta, 1-\delta)}}, \end{aligned} \tag{4.2}$$

where $\delta := (t'-t)/(b^{(i)} - a^{(i)})$. Therefore $m^{(i)}$ is absolutely continuous, since it is Lipshitz.

(d) *T is upper semi-continuous on (D_R, ρ_L), and concave.*

Remark 4.2. The integral in the definition of $T(\mathbf{f})$ is always finite whereas the term $D(\mu(\mathbf{f})\|\lambda_A)$ takes values in $[0, \infty]$. For $T(\mathbf{f})$ to be finite it is necessary that the component functions $f^{(i)}$ of $\mathbf{f}$ be absolutely continuous.

The above Theorem enables us to re-cast Theorem 3.2 in terms of our first notion of parametrized curve in §2, namely, in terms of random elements of the space D_L. $\mathbf{Z}_{\lceil n \cdot \rceil}$ is such a random element. Inverses are as in Remark 2.1.

COROLLARY 4.3. *Under the conditions of Theorem 3.2, for each $\varepsilon > 0$,*

$$P(\rho_u(\mathbf{Z}_{\lceil n \cdot \rceil}, \mathbf{m}^{\leftarrow}) > \varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \to 0 \qquad (n \to \infty).$$

Let $B^{\nearrow}$ be the set of non-decreasing right-continuous functions $\phi : [a^{(1)}, b^{(1)}] \to [a^{(2)}, b^{(2)}]$. This is a suitable space for a *non-parametrized* version of a curve joining the $\mathbf{Z}_j$ or $\mathbf{R}_j$, and for its limit(s). For $\phi \in B^{\nearrow}$ we have $\phi(x) = \int_{a^{(1)}}^x \dot{\phi}(u)\, du + \phi_s(x)$

where ϕ_s is singular. When ϕ_s is identically zero ϕ is absolutely continuous and we re-notate $\dot{\phi}$ as ϕ', its Radon-Nikodym derivative.

In [Deuschel & Zeitouni 1994] it is shown that under suitable conditions the limit non-parametrized curve for the $\mathbf{Z}_j$ is that (or those) ϕ maximizing the functional

$$J(\phi) := \int_{a^{(1)}}^{b^{(1)}} \sqrt{\dot{\phi}(x) h(x, \phi(x))}\, dx \qquad (\phi \in B^{\nearrow}).$$

Let us show that this functional is a projection of ours, so that the two solutions are consistent. The following is based on a communication from O. Zeitouni to one of us, which we acknowledge with gratitude.

For $\mathbf{f} \in D_R$ define $\phi_{\mathbf{f}}$ by

$$\phi_{\mathbf{f}}(x) := \begin{cases} \inf\{y : f^{(2)}(y) > f^{(1)}(x)\}, & \text{if } f^{(1)}(x) < 1, \\ b^{(2)}, & \text{if not,} \end{cases}$$

for $a^{(1)} \le x \le b^{(1)}$. One may check that $\phi_{\mathbf{f}} \in B^{\nearrow}$. As we shall establish in the proof of the next result, $f^{(2)}$ in cases of interest has an inverse $f^{(2)\leftarrow}$, so we are setting $\phi_{\mathbf{f}}(x) := f^{(2)\leftarrow}(f^{(1)}(x))$ here.

Proposition 4.4. *Assume H has density h which is such that $\log h$ is bounded on A. If $\mathbf{m}$ maximizes T over D_R then $\phi_{\mathbf{m}}$ maximizes J over $B^{\nearrow}$. Conversely, if ψ maximizes J over $B^{\nearrow}$ then it is absolutely continuous, and furthermore $\mathbf{f}_\psi$, given by*

$$f_\psi^{(1)\prime}(x) := \frac{\sqrt{h(x, \psi(x))\psi'(x)}}{J(\psi)}, \quad f_\psi^{(2)\prime}(y) := \frac{1}{J(\psi)} \sqrt{\frac{h(\psi^{\leftarrow}(y), y)}{\psi'(\psi^{\leftarrow}(y))}}, \tag{4.3}$$

maximizes T over D_R.

Remark 4.5. This result applies even in the Deuschel-Zeitouni cases of non-unique limits. That is, if there are several $\mathbf{m}$ that maximize T then each corresponding $\phi_{\mathbf{m}}$ maximizes J, while if there are several ψ that maximize J then each corresponding $\mathbf{f}_\psi$ maximizes T.

Proofs.

Lemma 4.6. *Let ν be a probability measure with support contained in a bounded interval $I := (a, b]$ in $\mathbb{R}$. Fix $k \in \mathbb{N}$, set $t_j := a + j(b-a)/k$ for $j = 0, 1, \ldots, k$, and consider a k^{th} approximation ν_k to ν, defined by*

$$d\nu_k(x) := \sum_{j=1}^{k} \mathbf{1}\{t_{j-1} < x \le t_j\} \frac{\nu(t_{j-1}, t_j]}{t_j - t_{j-1}}\, dx.$$

Then $D(\nu_k \| \lambda_I) \to D(\nu \| \lambda_I)$ as $k \to \infty$, where λ_I is the uniform probability law on I.

PROOF: Note that ν_k converges to ν narrowly. Since $D(\,\cdot\,\|\lambda_I)$ is lower semi-continuous with respect to the narrow topology (cf. e.g. [Deuschel & Stroock 1989, p. 69]), we have $D(\nu\|\lambda_I) \le \liminf D(\nu_k\|\lambda_I)$. We complete the proof by showing that $D(\nu_k\|\lambda_I) \le D(\nu\|\lambda_I)$ for each k. We may assume $D(\nu\|\lambda_I) < \infty$, so ν is absolutely continuous with density f, say. Then, with $g(x) := x\log x$,

$$\begin{aligned} D(\nu_k\|\lambda_I) &= \sum_{j=1}^{k}(t_j - t_{j-1})\frac{\nu(t_{j-1},t_j]}{t_j - t_{j-1}}\log\frac{\nu(t_{j-1},t_j](b-a)}{t_j - t_{j-1}} \\ &= \sum_{j=1}^{k}(t_j - t_{j-1})\, g\!\left(\frac{\int_{t_{j-1}}^{t_j} f(x)\,dx}{t_j - t_{j-1}}\right) + \nu(a,b]\log(b-a) \\ &\le \sum_{j=1}^{k}(t_j - t_{j-1})\frac{1}{t_j - t_{j-1}}\int_{t_{j-1}}^{t_j} g(f(x))\,dx + \nu(a,b]\log(b-a) \end{aligned}$$

(by Jensen's inequality)

$$= \int_a^b f(x)\log\big((b-a)f(x)\big)\,dx.$$

The right-hand side is $D(\nu\|\lambda_I)$. $\square$

LEMMA 4.7. *For each* $\mathbf{f} \in D_R$,

$$\overline{T}_k(\mathbf{d}_k(\mathbf{f})) \to T(\mathbf{f}) \qquad (k \to \infty). \tag{4.4}$$

Further, given $\varepsilon > 0$ *there exists* $k = k_0(\varepsilon)$ *such that for all* $k \ge k_0$,

$$T(\mathbf{c}_k(\mathbf{p})) - \varepsilon \le \underline{T}_k(\mathbf{p}) \le \overline{T}_k(\mathbf{p}) \le T(\mathbf{c}_k(\mathbf{p})) + \varepsilon \qquad (\mathbf{p} \in \Delta_k^2). \tag{4.5}$$

PROOF: $\log h$ is uniformly continuous on A, so we may choose k_0 such that $\log \overline{h}_k(\mathbf{z}) - \log \underline{h}_k(\mathbf{z}) \le \varepsilon$ for all $\mathbf{z} \in A$ and $k \ge k_0$. By Lemma 3.5,

$$0 \le \overline{S}_k(\mathbf{d}_k(\mathbf{f})) - \int_0^1 \log\frac{h(\mathbf{f}^{\leftarrow}(p))}{H(A)}\,dp \le \varepsilon \qquad (\mathbf{f} \in D_R,\ k \ge k_0). \tag{4.6}$$

Now it is readily checked that

$$U_k(\mathbf{p}) = D(\mu(\mathbf{c}_k(\mathbf{p}))\|\lambda_A) \qquad (\mathbf{p} \in \Delta_k^2). \tag{4.7}$$

For any $\mathbf{f} \in D_R$, the measure $\mu_k(\mathbf{f}) := \mu\big(\mathbf{c}_k(\mathbf{d}_k(\mathbf{f}))\big)$ is a smoothed version of $\mu(\mathbf{f})$, the smoothing being that of Lemma 4.6. From (4.7) we have $U_k(\mathbf{d}_k(\mathbf{f})) = D(\mu_k(\mathbf{f})\|\lambda_A)$, and so by Lemma 4.6,

$$U_k(\mathbf{d}_k(\mathbf{f})) \to D(\mu(\mathbf{f})\|\lambda_A) \qquad (k \to \infty).$$

With (4.6) this gives (4.4).

With $k_0(\varepsilon)$ as above, pick $k \geq k_0$ and $\mathbf{p} \in \Delta_k^2$, and set $\mathbf{f} := \mathbf{c}_k(\mathbf{p})$. Then $\mathbf{d}_k(\mathbf{f}) = \mathbf{p}$ by (3.17), so the right-hand inequality of (4.5) follows from (4.6) and (4.7). The left-hand inequality may be established similarly. $\square$

Proof of Theorem 4.1: We first prove (a)–(d).

(a) We first show that

$$\rho_L(\mathbf{c}_k(\mathbf{q}_k), \mathbf{m}) \to 0 \qquad (k \to \infty), \tag{4.8}$$

where $\mathbf{q}_k$ is as in Lemma 3.10. Fix $\varepsilon > 0$. In (3.24) we proved that for each sufficiently large k,

$$P(\rho_L(\mathbf{M}_n, \mathbf{c}_k(\mathbf{q}_k)) \leq \tfrac{1}{2}\varepsilon \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \to 1 \qquad (n \to \infty).$$

Fix upon any one of these k. Pick a value of n so that the above conditional probability exceeds $\frac{1}{2}$, and so that also $P(\rho_L(\mathbf{M}_n, \mathbf{m}) \leq \varepsilon/2 \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n)$ (which we know converges to 1, by Theorem 3.2) exceeds $\frac{1}{2}$. There must thus exist an element ω of the event $\{\mathbf{Z}_1 < \cdots < \mathbf{Z}_n\}$ such that both $\rho_L(\mathbf{M}_n(\omega), \mathbf{c}_k(\mathbf{q}_k)) \leq \varepsilon/2$ and $\rho_L(\mathbf{M}_n(\omega), \mathbf{m}) \leq \varepsilon/2$. Therefore $\rho_L(\mathbf{c}_k(\mathbf{q}_k), \mathbf{m}) \leq \varepsilon$, and (4.8) is proved.

We prove $\mathbf{m}$ is continuous. Choose $\varepsilon > 0$, then make k so large that $\rho_L(\mathbf{c}_k(\mathbf{q}_k), \mathbf{m}) < \varepsilon$. This implies that for $i = 1, 2$,

$$m^{(i)}(x+\varepsilon) - m^{(i)}(x) \leq c_k^{(i)}(\mathbf{q}_k^{(i)})(x+2\varepsilon) - c_k^{(i)}(\mathbf{q}_k^{(i)})(x-\varepsilon) + 2\varepsilon.$$

The difference between the values of $c_k^{(i)}(\mathbf{q}_k^{(i)})$ on the right is at most the sum of the numbers $q_{kj}^{(i)}$ over $\lceil 3\varepsilon k/(b^{(i)} - a^{(i)}) \rceil + 1$ consecutive values of j. From (3.7) it follows that this sum is at most a bounded multiple of ε. Continuity of $m^{(i)}$ follows.

With $\mathbf{m}$ known to be continuous, (4.8) becomes equivalent to

$$\rho_u(\mathbf{c}_k(\mathbf{q}_k), \mathbf{m}) \to 0 \qquad (k \to \infty), \tag{4.9}$$

and we also deduce (a) from (3.1).

(b) Continuity has been proved above. Since for $i = 1, 2$ and all k, $\mathbf{c}_k(\mathbf{q}_k)^{(i)}(a^{(i)}) = 0$ and $\mathbf{c}_k(\mathbf{q}_k)^{(i)}(b^{(i)}) = 1$ by construction, the values of $\mathbf{m}$ at the end-points follow from (4.9).

(c) By (4.9),

$$\begin{aligned} m^{(i)}(t') - m^{(i)}(t) &= \lim_{k\to\infty} (\mathbf{c}_k(\mathbf{q}_k)^{(i)}(t') - \mathbf{c}_k(\mathbf{q}_k)^{(i)}(t)) \\ &= \lim_{k\to\infty} \sum q_{kj}^{(i)}, \end{aligned}$$

where the sum is over those integers j satisfying

$$\frac{t - a^{(i)}}{b^{(i)} - a^{(i)}} k \leq j < \frac{t' - a^{(i)}}{b^{(i)} - a^{(i)}} k.$$

This is a sum of the $q_{kj}^{(i)}$ over $m = k\delta \pm 1$ consecutive indices j, so must satisfy the (3.7) bounds. As $k \to \infty$, $m/k \to \delta$, so the bounds converge to those in (4.2), which is thus proved.

(d) Since $\mathbf{d}_k$ is a linear map and $\overline{T}_k = \overline{S}_k - U_k$ is concave, the composition $\overline{T}_k \circ \mathbf{d}_k = \overline{S}_k \circ \mathbf{d}_k - U_k \circ \mathbf{d}_k$ is concave. By (4.4), T is concave.

We now show that the map $\mathbf{f} \mapsto \int_0^1 \log h(\mathbf{f}^{\leftarrow}(p))\,dp$ is continuous on (D_R, ρ_L). If $\rho_L(\mathbf{f}_n, \mathbf{f}) \to 0$, where $\mathbf{f}_n$, $\mathbf{f} \in D_R$, each component $f_n^{(i)}$ converges to $f^{(i)}$ in the Skorokhod M_1 topology. Inversion is a continuous operation under this topology (cf. [Whitt 1980, §7]), so $f_n^{(i)\leftarrow} \to f^{(i)\leftarrow}$ in the M_1 sense. Since the map $\mathbf{g} \mapsto \int_0^1 \log h(\mathbf{g}(p))\,dp$ on D_L is M_1–continuous, we have the continuity of the map we want.

As noted earlier, $D(\,\cdot\,\|\lambda_A)$ is lower semi-continuous on the space M_A^1 of probability measures on A, under the topology of narrow convergence. Since (D_R, ρ_L) is equivalent to the subset of product measures in M_A^1, it follows that $D(\mu(\cdot)\|\lambda_A)$ is lower semi-continuous on (D_R, ρ_L). With the continuity proved in the previous paragraph we thus have (d).

It remains to prove that $\mathbf{m}$ is the unique element of D_R that maximizes T. First, $T(\mathbf{m}) \geq \limsup_{k\to\infty} T(\mathbf{c}_k(\mathbf{q}_k))$, by (4.8) and upper semi-continuity of T .

Since T is concave, there is an $\mathbf{f}_0 \in D_R$ at which T attains its supremum. Choose $\varepsilon > 0$ and fix k large enough, as by (4.4) we may, so that

$$\overline{T}_k(\mathbf{d}_k(\mathbf{f}_0)) > \sup T - \tfrac{1}{3}\varepsilon. \tag{4.10}$$

Also make k so large that $T(\mathbf{m}) > T(\mathbf{c}_k(\mathbf{q}_k)) - \varepsilon/3$, by the above, and to be at least $k_0(\varepsilon/3)$ in Lemma 4.7. The latter two considerations give that $T(\mathbf{m}) > \overline{T}_k(\mathbf{q}_k) - \frac{2}{3}\varepsilon$. Since $\mathbf{q}_k$ maximizes $\overline{T}_k$ it follows that $T(\mathbf{m}) > \overline{T}_k(\mathbf{d}_k(\mathbf{f}_0)) - \frac{2}{3}\varepsilon$. With (4.10) this gives $T(\mathbf{m}) > \sup T - \varepsilon$, and so $\mathbf{m}$ maximizes T. We have also established

$$T(\mathbf{m}) = \lim_{k\to\infty} \overline{T}_k(\mathbf{q}_k). \tag{4.11}$$

To prove uniqueness, suppose both $\mathbf{m}$ and $\mathbf{m}_0$ maximize T. Then all convex combinations of them do so too, as T is concave. Since both $\int_0^1 \log h(\cdot^{\leftarrow}(p))\,dp$ and $-D(\mu(\cdot)\|\lambda_A)$ are concave, it follows that $D(\mu(\delta\mathbf{m}+(1-\delta)\mathbf{m}_0)\|\lambda_A) = \delta D(\mu(\mathbf{m})\|\lambda_A) + (1-\delta)D(\mu(\mathbf{m}_0)\|\lambda_A)$ for all $0 \leq \delta \leq 1$. The right-hand side is

$$\delta \int_A d\lambda_A \frac{d\mu(\mathbf{m})}{d\lambda_A} \log \frac{d\mu(\mathbf{m})}{d\lambda_A} + (1-\delta) \int_A d\lambda_A \frac{d\mu(\mathbf{m}_0)}{d\lambda_A} \log \frac{d\mu(\mathbf{m}_0)}{d\lambda_A}.$$

Since $\mu(\delta\mathbf{m} + (1-\delta)\mathbf{m}_0) = \delta\mu(\mathbf{m}) + (1-\delta)\mu(\mathbf{m}_0)$, the left-hand side is

$$\int_A d\lambda_A \left(\delta \frac{d\mu(\mathbf{m})}{d\lambda_A} + (1-\delta)\frac{d\mu(\mathbf{m}_0)}{d\lambda_A}\right) \log\left(\delta \frac{d\mu(\mathbf{m})}{d\lambda_A} + (1-\delta)\frac{d\mu(\mathbf{m}_0)}{d\lambda_A}\right).$$

Because of the strict convexity of the function $g(x) = x\log x$, equality of the two sides forces the Radon-Nikodym derivatives $d\mu(\mathbf{m})/d\lambda_A$ and $d\mu(\mathbf{m}_0)/d\lambda_A$ to coincide a.e., and that implies $\mathbf{m} = \mathbf{m}_0$. Thus the maximizing $\mathbf{m}$ is unique. $\square$

Proof of Corollary 4.3: From Theorem 4.1(b) we know $m^{(i)}(a^{(i)}) = 0$ and $m^{(i)}(b^{(i)}) = 1$, while from (e) of the same result $m^{(i)}$ is absolutely continuous (since $T(\mathbf{m}) > -\infty$). Indeed, (c) yields that $m^{(i)\prime}(t) \ge c$ for $0 < t < 1$, where $c := e^{-6C_h} > 0$. Suppose that $\mathbf{f} \in D_R$ has $\delta := \rho_u(\mathbf{f}, \mathbf{m}) < 1$, so that $|f^{(i)}(t) - m^{(i)}(t)| \le \delta$ for $a^{(i)} \le t \le b^{(i)}$. Then it is elementary that $|f^{(i)\leftarrow}(p) - m^{(i)\leftarrow}(p)| \le \delta/c$ for $0 \le p \le 1$, so that $\rho_u(\mathbf{f}^{\leftarrow}, \mathbf{m}^{\leftarrow}) \le \delta/c$. On the event $\mathbf{Z}_1 < \cdots < \mathbf{Z}_n$, $\mathbf{M}_n^{\leftarrow}$ coincides with $\mathbf{Z}_{\lceil n\cdot\rceil}$, so we conclude the result by applying the above to (4.1). □

Proof of Proposition 4.4 (based on a communication from O. Zeitouni): For $\mathbf{f} \in D_R$ such that $T(\mathbf{f}) > -\infty$, $f^{(1)}$ and $f^{(2)}$ are absolutely continuous, and

$$\begin{aligned} T(\mathbf{f}) = \int_0^1 dp\, \log h\big(f^{(1)\leftarrow}(p), f^{(2)\leftarrow}(p)\big) - \int_{a^{(1)}}^{b^{(1)}} dx\, f^{(1)\prime}(x) \log f^{(1)\prime}(x) \\ - \int_{a^{(2)}}^{b^{(2)}} dy\, f^{(2)\prime}(y) \log f^{(2)\prime}(y) - \log H(A)\lambda(A). \end{aligned}$$

Any candidate for maximizing T must also have $f^{(i)\prime}$ strictly positive over $(a^{(i)}, b^{(i)})$; this may be proved by an argument involving directional derivatives, similar to that in the proof of Lemma 3.10 but in the space D_R in place of Δ_k^2, and again using the fact that the function $x \log x$ has infinite slope at $0+$. We may thus restrict consideration to such $\mathbf{f}$. Then $\phi = \phi_{\mathbf{f}}$ will be a strictly increasing differentiable function from $[a^{(1)}, b^{(1)}]$ onto $[a^{(2)}, b^{(2)}]$, given by $\phi(x) = f^{(2)\leftarrow}(f^{(1)}(x))$.

In the first integral in the above formula for $T(\mathbf{f})$ substitute $p = f^{(1)}(x)$, making it

$$\int_{a^{(1)}}^{b^{(1)}} dx\, f^{(1)\prime}(x) \log h(x, \phi(x)),$$

and in the third integral substitute $y = \phi(x)$, making it

$$\int_{a^{(1)}}^{b^{(1)}} dx\, \phi'(x) f^{(2)\prime}(\phi(x)) \log f^{(2)\prime}(\phi(x)) = \int_{a^{(1)}}^{b^{(1)}} dx\, f^{(1)\prime}(x) \log \frac{f^{(1)\prime}(x)}{\phi'(x)}.$$

So

$$\begin{aligned} T(\mathbf{f}) + \log H(A)\lambda(A) &= \int_{a^{(1)}}^{b^{(1)}} dx\, f^{(1)\prime}(x) \log \frac{h(x, \phi(x))\phi'(x)}{(f^{(1)\prime}(x))^2} \\ &= 2 \int_{a^{(1)}}^{b^{(1)}} dx\, f^{(1)\prime}(x) \log \frac{\sqrt{h(x, \phi(x))\phi'(x)}}{f^{(1)\prime}(x)}. \end{aligned}$$

We normalize the numerator of the fraction:

$$\begin{aligned} &\frac{1}{2}\big(T(\mathbf{f}) + \log H(A)\lambda(A)\big) \\ &\quad = \int_{a^{(1)}}^{b^{(1)}} dx\, f^{(1)\prime}(x) \log \frac{J(\phi)^{-1}\sqrt{h(x, \phi(x))\phi'(x)}}{f^{(1)\prime}(x)} + \log J(\phi). \end{aligned} \tag{4.12}$$

We apply the 'Gibbs' inequality, that for probability densities p, q on $\mathbb{R}$,

$$\int_{\{x: p(x)q(x)>0\}} p(x) \log \frac{q(x)}{p(x)}\, dx \leq 0,$$

with equality if and only if $p = q$ a.e. (cf. e.g. [Goldie & Pinch 1991, Lemma 4.5.1]). Since both $f^{(1)\prime}$ and $J(\phi)^{-1}\sqrt{h(x,\phi(x))\phi'(x)}$ are probability densities the inequality says the integral in (4.12) is non-positive, and equals zero if and only if the two densities coincide a.e. The supremum of $T(\mathbf{f})$ over all $\mathbf{f}$ consistent with the fixed $\phi \in B^{\nearrow}$ is thus

$$\sup_{\{\mathbf{f}: \phi_{\mathbf{f}} = \phi\}} T(\mathbf{f}) = 2 \log J(\phi) - \log H(A)\lambda(A).$$

Thus T is maximized if and only if J is maximized and the optimum ϕ and $\mathbf{f}$ are connected by $\phi = \phi_{\mathbf{f}}$. We have also shown above that for the optimal $\mathbf{f}$ and ϕ, namely $\mathbf{m}$ and ψ, $m^{(1)}$ is given by the left-hand formula in (4.3), and a further calculation using $m^{(2)}(x) = m^{(1)}(\psi^{\leftarrow}(x))$ then yields the formula for $m^{(2)}$. $\square$

5. Large deviations.

Theorem 5.1. *Under the conditions of Theorem 3.2,*

$$P(\mathbf{Z}_1 < \cdots < \mathbf{Z}_n) = e^{n(-2\log n + 2 + \log \lambda(A) + T(\mathbf{m}) + o(1))} \qquad (n \to \infty) \tag{5.1}$$

and

$$P(\mathbf{Z}_1, \ldots, \mathbf{Z}_n \text{ can be ordered}) = e^{n(-\log n + 1 + \log \lambda(A) + T(\mathbf{m}) + o(1))} \qquad (n \to \infty). \tag{5.2}$$

Further, for every $\Gamma \in \mathcal{B}(D_R)$, as $n \to \infty$,

$$\begin{aligned} \exp\Big(n\Big(\sup_{\Gamma^\circ} T + o(1)\Big)\Big) &\leq e^{n(2\log n - 2 - \log \lambda(A))} P(\mathbf{M}_n \in \Gamma,\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ &\leq \exp\Big(n\Big(\sup_{\bar{\Gamma}} T + o(1)\Big)\Big), \end{aligned} \tag{5.3}$$

with closure and interior under the ρ_L metric. The latter implies that

$$\begin{aligned} -\inf_{\Gamma^\circ} I &\leq \liminf_{n\to\infty} \frac{1}{n} \log P(\mathbf{M}_n \in \Gamma \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P(\mathbf{M}_n \in \Gamma \mid \mathbf{Z}_1 < \cdots < \mathbf{Z}_n) \leq -\inf_{\bar{\Gamma}} I, \end{aligned}$$

where $I : D_R \to [0, \infty]$ is defined by $I(\mathbf{f}) := T(\mathbf{m}) - T(\mathbf{f})$, and is a 'good, convex rate function' in the terminology of [Deuschel & Stroock 1989].

PROOF: We write $P_n(A)$ for $P(A, \mathbf{Z}_1 < \cdots < \mathbf{Z}_n)$ throughout this proof.

First note that (5.1) will follow from (5.3), on taking Γ to be the whole space D_R. Then (5.2) follows since its left-hand side is that of (5.1), times $n!$.

Second, I is a 'good, convex rate-function' because by Theorem 4.1(d) it is lower semi-continuous and convex, on a compact space.

So the only task is to prove (5.3). Let $\mathbf{M}_{nk} := \mathbf{c}_k(\mathbf{d}_k(\mathbf{M}_n))$. We first fix upon a Borel set Γ in (D_R, ρ_L) and prove that, given $\varepsilon > 0$, for all sufficiently large k and for all n exceeding some $n_0(k)$ we have

$$\begin{aligned}\exp\bigl(n\bigl(&-2\log n + 2 + \log\lambda(A) + \sup_{\Gamma^\circ} T - \varepsilon\bigr)\bigr)\\ &\le P_n(\mathbf{M}_{nk} \in \Gamma)\\ &\le \exp\bigl(n\bigl(-2\log n + 2 + \log\lambda(A) + \sup_{\bar\Gamma} T + \varepsilon\bigr)\bigr).\end{aligned} \tag{5.4}$$

For the upper bound, note that T is finite at each point $\mathbf{c}_k(\mathbf{p})$, for $\mathbf{p} \in \Delta_k^2$, so the bound holds trivially if T is identically $-\infty$ on $\bar\Gamma$, both sides being 0, since for no $\mathbf{p}$ is it then the case that $\mathbf{c}_k(\mathbf{p}) \in \bar\Gamma$. Otherwise, fix $k \ge k_0(\varepsilon/2)$ in Lemma 4.7, and calculate

$$\begin{aligned}P_n(\mathbf{M}_{nk} \in \Gamma) &= \sum_{\{\mathbf{p}\in\Delta_{kn}^2 : \mathbf{c}_k(\mathbf{p})\in\Gamma\}} P_n(\mathbf{d}_k(\mathbf{M}_n) = \mathbf{p})\\ &\le \sum_{\{\mathbf{p}\in\Delta_{kn}^2 : \mathbf{c}_k(\mathbf{p})\in\Gamma\}} \exp\bigl(n(-2\log n + 2 + \log\lambda(A) + \overline{T}_k(\mathbf{p}))\bigr)\\ &\le \sum_{\{\mathbf{p}\in\Delta_{kn}^2 : \mathbf{c}_k(\mathbf{p})\in\Gamma\}} \exp\bigl(n(-2\log n + 2 + \log\lambda(A) + T(\mathbf{c}_k(\mathbf{p})) + \tfrac12\varepsilon)\bigr)\\ &\le n^{2k}\exp\Bigl(n\Bigl(-2\log n + 2 + \log\lambda(A) + \sup_{\bar\Gamma} T + \tfrac12\varepsilon\Bigr)\Bigr),\end{aligned}$$

where we used (3.18) for the first inequality and (4.5) for the second. The bound follows, for all sufficiently large n.

The lower bound holds trivially if $\sup_{\Gamma^\circ} T = -\infty$, so assume otherwise and pick $\mathbf{f}_\varepsilon \in \Gamma^\circ$ so that $T(\mathbf{f}_\varepsilon) > \sup_{\Gamma^\circ} T - \frac18\varepsilon$. Let $\delta > 0$ be such that all $\mathbf{f} \in D_R$ with $\rho_L(\mathbf{f}, \mathbf{f}_\varepsilon) < \delta$ are in Γ.

Now because $T(\mathbf{f}_\varepsilon)$ is finite, $\mathbf{f}_\varepsilon$ is in the set C of continuous elements of D_R, so $|\mathbf{d}_k(\mathbf{f}_\varepsilon)| < \frac14\delta$ for all $k \ge k_1$, say (where again the $|\cdot|$ notation is that of Lemma 3.14). Fix upon such a k that also exceeds $k_0(\varepsilon/16)$ in Lemma 4.7, and is large enough for

$$\overline{T}_k(\mathbf{d}_k(\mathbf{f}_\varepsilon)) > T(\mathbf{f}_\varepsilon) - \tfrac18\varepsilon, \tag{5.5}$$

as (4.4) ensures is possible. Write $\mathbf{q} := \mathbf{d}_k(\mathbf{f}_\varepsilon)$.

As the restriction of $\overline{T}_k$ to Δ_k^2 is continuous we may find $\eta > 0$ so small that $\overline{T}_k(\mathbf{p}) > \overline{T}_k(\mathbf{q}) - \frac18\varepsilon$ for all $\mathbf{p} \in B_\eta(\mathbf{q})$. We also make $\eta < \frac14\delta$. Then for all $\mathbf{p} \in B_\eta(\mathbf{q})$, $|\mathbf{p}| \le \|\mathbf{p} - \mathbf{q}\| + |\mathbf{q}| < \frac12\delta$ so that by Lemma 3.14,

$$\rho_L(\mathbf{c}_k(\mathbf{p}), \mathbf{f}_\varepsilon) \le \|\mathbf{p} - \mathbf{q}\| + |\mathbf{p}| + |\mathbf{q}| < \tfrac14\delta + \tfrac12\delta + \tfrac14\delta = \delta.$$

Thus $\mathbf{c}_k(\mathbf{p}) \in \Gamma$. Also

$$\begin{aligned}\underline{T}_k(\mathbf{p}) &> \overline{T}_k(\mathbf{p}) - \tfrac{1}{8}\varepsilon \qquad \text{(by (4.5))}\\ &> \overline{T}_k(\mathbf{q}) - \tfrac{1}{4}\varepsilon\\ &> T(\mathbf{f}_\varepsilon) - \tfrac{3}{8}\varepsilon \qquad \text{(by (5.5))}\\ &> \sup_{\Gamma^\circ} T - \tfrac{1}{2}\varepsilon.\end{aligned}$$

Finally,

$$\begin{aligned}&P_n(\mathbf{M}_{nk} \in \Gamma)\\ &\quad = \sum_{\{\mathbf{p}\in\Delta^2_{kn}:\mathbf{c}_k(\mathbf{p})\in\Gamma\}} P_n(\mathbf{d}_k(\mathbf{M}_n) = \mathbf{p})\\ &\quad \geq \sum_{\mathbf{p}\in\Delta^2_{kn}\cap B_\eta(\mathbf{q})} \exp\bigl(n\bigl(-2\log n + 2 + \log\lambda(A) + \underline{T}_k(\mathbf{p})\bigr) - k\log(2\pi e^2 n)\bigr)\end{aligned}$$

(by (3.18))

$$\geq \sum_{\mathbf{p}\in\Delta^2_{kn}\cap B_\eta(\mathbf{q})} \exp\bigl(n\bigl(-2\log n + 2 + \log\lambda(A) + \underline{T}_k(\mathbf{p}) - \tfrac{1}{2}\varepsilon\bigr)\bigr)$$

(for large n)

$$\geq \sum_{\mathbf{p}\in\Delta^2_{kn}\cap B_\eta(\mathbf{q})} \exp\Bigl(n\Bigl(-2\log n + 2 + \log\lambda(A) + \sup_{\Gamma^\circ} T - \varepsilon\Bigr)\Bigr),$$

and the sum is over a non-empty set if n is large enough. The lower bound in (5.4) follows.

We next establish that for all $\delta > 0$,

$$\lim_{k\to\infty}\limsup_{n\to\infty} n^{-1}\bigl(\log P_n(\rho_L(\mathbf{M}_{nk}, \mathbf{M}_n) > \delta) + 2n\log n\bigr) = -\infty. \tag{5.6}$$

Let $A_\delta := \{\mathbf{p} \in \Delta^2_k : p^{(i)}_j > \delta \text{ for some } i, j\}$. Now if $(p_1, \ldots, p_k)$ is an element of Δ_k such that one of its components p_i has value $\delta' \geq \delta$ then the elementary 'Fano inequality' of information theory (see, e.g., [Goldie & Pinch 1991, Theorem 2.9.3]) gives that

$$\begin{aligned}-\sum_1^k p_j\log p_j &\leq \log 2 + (1-\delta')\log(k-1)\\ &\leq \log 2 + (1-\delta)\log(k-1).\end{aligned}$$

Without this restriction the entropy $-\sum_1^k p_j \log p_j$ has upper bound $\log k$. Thus if $\mathbf{p} \in A_\delta$ then

$$\begin{aligned}
-U_k(\mathbf{p}) &= -2\log k - \sum_{i=1}^{2}\sum_{j=1}^{k} p_j^{(i)} \log p_j^{(i)} \\
&\le -2\log k + \log k + \log 2 + (1-\delta)\log(k-1) \\
&\le \log 2 + (1-\delta)\log\bigl(1-k^{-1}\bigr) - \delta \log k \\
&\le \log 2 - \delta \log k.
\end{aligned}$$

Since

$$P_n(\rho_L(\mathbf{M}_{nk}, \mathbf{M}_n) > \delta) \le \sum_{\mathbf{p} \in \Delta_{kn}^2 \cap A_\delta} P_n(\mathbf{d}_k(\mathbf{M}_n) = \mathbf{p}),$$

we obtain from (3.18) that

$$\begin{aligned}
e^{2n\log n} P_n(\rho_L(\mathbf{M}_{nk}, \mathbf{M}_n) > \delta) &\le \sum_{\mathbf{p} \in \Delta_{kn}^2 \cap A_\delta} \exp\bigl(n(\log\lambda(A) + \overline{T}_k(\mathbf{p}))\bigr) \\
&\le \sum_{\mathbf{p} \in \Delta_{kn}^2 \cap A_\delta} \exp\bigl(n(\log\lambda(A) + C_h - U_k(\mathbf{p}))\bigr) \\
&\le n^{2k} e^{n(\log 2\lambda(A) + C_h - \delta \log k)},
\end{aligned}$$

and (5.6) is proved.

Let F be a closed subset of (D_R, ρ_L) and let F_δ denote the open δ–neighbourhood around it. Then for any $\delta > 0$ and $k \ge 1$,

$$P_n(\mathbf{M}_n \in F) \le P_n(\mathbf{M}_{nk} \in \overline{F_\delta}) + P_n(\rho_L(\mathbf{M}_{nk}, \mathbf{M}_n) > \delta).$$

Hence, by (5.4) and (5.6),

$$\limsup_{n\to\infty} \left(\frac{1}{n}\log P_n(\mathbf{M}_n \in F) + 2\log n - 2 - \log\lambda(A)\right) \le \sup_{\overline{F_\delta}} T \tag{5.7}$$

for every $\delta > 0$. It is well known that $\sup_{\overline{F_\delta}} T \to \sup_F T$ as $\delta \downarrow 0$, at least if the assertion is put in terms of the rate function I; cf. [Deuschel & Stroock 1989, Lemma 2.1.2]. So the upper bound in (5.3) is proved.

Let G be open in (D_R, ρ_L), pick $\mathbf{f} \in G$ and make $\delta > 0$ so small that the open 2δ–ball around $\mathbf{f}$, $N_{2\delta}(\mathbf{f})$, is contained in G. Then for each $k \ge 1$,

$$P_n(\mathbf{M}_{nk} \in N_\delta(\mathbf{f})) \le P_n(\mathbf{M}_n \in G) + P(\rho_L(\mathbf{M}_{nk}, \mathbf{M}_n) \ge \delta).$$

By (5.4) and (5.6) it follows that

$$\liminf_{n\to\infty} \left(\frac{1}{n}\log P_n(\mathbf{M}_n \in G) + 2\log n - 2 - \log\lambda(A)\right) \ge \sup_{N_\delta(\mathbf{f})} T,$$

whence the lower bound in (5.3). $\square$

6. Application to records.

THEOREM 6.1. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be iid with distribution F having continuous density f on A. Set $h(\mathbf{x}) := f(\mathbf{x})/(1 - F(\mathbf{x}))$ and suppose further that*

(a) $F(\mathbf{b}) < 1$;
(b) $\log f$ *is bounded on* A;
(c) $\log h$ *is L-superadditive on* A.

Define $\mathbf{L}$ *as in §2. Then there exists a non-random* $\mathbf{m} \in D_R$ *such that for every* $\varepsilon > 0$,

$$P(\rho_u(\mathbf{L}, \mathbf{m}) > \varepsilon \mid N_A = n) \to 0 \qquad (n \to \infty) \tag{6.1}$$

and

$$P(\rho_u(\mathbf{R}_{\lceil n\cdot\rceil}, \mathbf{m}^{\leftarrow}) > \varepsilon \mid N_A = n) \to 0 \qquad (n \to \infty). \tag{6.2}$$

Further, $\mathbf{m}$ *is the unique element of* D_R *that maximizes* T*, and the other conclusions (b)–(d) of Theorem 4.1 also hold.*

COROLLARY 6.2. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be iid with independent continuous components: $\mathbf{X}_1 = (X_1^{(1)}, X_1^{(2)})$ and*

$$P(X_1^{(1)} \in dx^{(1)},\ X_1^{(2)} \in dx^{(2)}) = f_1(x^{(1)}) f_2(x^{(2)})\, dx^{(1)} dx^{(2)},$$

where each density f_i is continuous and has bounded logarithm on $[a^{(i)}, b^{(i)}]$. Suppose also that

$$P(X_1^{(1)} \le b^{(1)}) P(X_1^{(2)} \le b^{(2)}) < 1.$$

Then the result of Theorem 6.1 holds.

THEOREM 6.3. *Under the conditions of Theorem 6.1, and in particular under the conditions of Corollary 6.2,*

$$P(N_A = n) = \frac{(\lambda(A) e^{T(\mathbf{m})} + o(1))^n}{(n!)^2} \qquad (n \to \infty), \tag{6.3}$$

and likewise for $P(N_A \ge n)$.

Further, for every $\Gamma \in \mathcal{B}(D_R)$, as $n \to \infty$,

$$\begin{aligned} \exp\Big(n\Big(\sup_{\Gamma^\circ} T + o(1)\Big)\Big) &\le e^{n(2\log n - 2 - \log\lambda(A))} P(\mathbf{L} \in \Gamma,\ N_A = n) \\ &\le \exp\Big(n\Big(\sup_{\bar{\Gamma}} T + o(1)\Big)\Big), \end{aligned} \tag{6.4}$$

with closure and interior under the ρ_L metric. The latter implies that, with I as in Theorem 5.1,

$$\begin{aligned} -\inf_{\Gamma^\circ} I &\le \liminf_{n\to\infty} \frac{1}{n} \log P(\mathbf{L} \in \Gamma \mid N_A = n) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log P(\mathbf{L} \in \Gamma \mid N_A = n) \le -\inf_{\bar{\Gamma}} I. \end{aligned}$$

Proofs.

PROOF OF THEOREM 6.1: (a) and (b) imply $\log h$ is bounded on A. With (c), the conditions of Theorem 3.2 are thus satisfied. So the conclusions of Theorems 3.2 and 4.1 hold. The latter gives that the components $m^{(i)}$ of $\mathbf{m}$ are strictly increasing, and that fact together with the conclusion of Theorem 3.2 yields (2.5). So by Theorem 2.2 we conclude (2.7), and the properties of $\mathbf{m}$ then allow us to deduce (6.1) and (6.2), exactly as in Theorem 4.1(a) and Corollary 4.3. □

PROOF OF COROLLARY 6.2: Since

$$h(x^{(1)}, x^{(2)}) = \frac{f_1(x^{(1)})f_2(x^{(2)})}{1 - F_1(x^{(1)})F_2(x^{(2)})},$$

$\log h$ is L-superadditive if and only if $-\log(1 - F_1(x^{(1)})F_2(x^{(2)}))$ is L-superadditive, and the latter is so by Proposition 3.1 because

$$\frac{\partial^2 - \log(1 - F(x^{(1)})F(x^{(2)}))}{\partial x^{(2)} \partial x^{(1)}} = \frac{f_1(x^{(1)})f_2(x^{(2)})}{(1 - F_1(x^{(1)})F_2(x^{(2)}))^2} > 0$$

on $A°$. The conditions of Theorem 6.1 are thus satisfied, whence the result. □

LEMMA 6.4. *Under the conditions of Theorem 6.1 there exists a constant C_Q such that $0 \geq \log Q(B) \geq -C_Q H(B)$, for all Borel subsets B of A.*

PROOF: By its definition, Q is non-increasing: $Q(B) \geq Q(C)$ if $B \subseteq C$. So for any Borel set $B \subseteq A$,

$$Q(B) \geq Q(\{\mathbf{x} : \mathbf{x} \leq \mathbf{b}\}) = 1 - F(\mathbf{b}) > 0,$$

the equality being by [Goldie & Resnick 1989, Corollary 2.5]. Put $C_Q := -(F(\mathbf{b}))^{-1} \log(1 - F(\mathbf{b})) > 0$, then $\log x \geq -C_Q(1 - x)$ for $1 - F(\mathbf{b}) \leq x \leq 1$, so that from the above, $0 \geq \log Q(B) \geq -C_Q(1 - Q(B))$, and the result follows via (2.8). □

PROOF OF THEOREM 6.3: (6.3) follows from taking $\Gamma := D_R$ in (6.4). Also the large-deviation bounds follow from (6.4). To prove (6.4), note that

$$P(\mathbf{L} \in \Gamma,\ N_A = n) = \int_{A^{(n)}} \mathbf{1}\{\mathbf{e}(\mathbf{z}_1, \ldots, \mathbf{z}_n) \in \Gamma\} H(d\mathbf{z}_1) \cdots H(d\mathbf{z}_n) \prod_{j=0}^{n} Q((\mathbf{z}_j, \mathbf{z}_{j+1})),$$

with $\mathbf{z}_0 := \mathbf{a}$, $\mathbf{z}_{n+1} := \mathbf{b}$, and that the right-hand side becomes $P(\mathbf{M}_n \in \Gamma,\ \mathbf{Z}_1 < \cdots < \mathbf{Z}_n)$ if we omit the product at the end. Now by Lemma 6.4,

$$\begin{aligned} 1 \geq \prod_{j=0}^{n} Q((\mathbf{z}_j, \mathbf{z}_{j+1})) &\geq \exp\Big(-C_Q \sum_{j=0}^{n} H((\mathbf{z}_j, \mathbf{z}_{j+1}))\Big) \\ &= \exp\Big(-C_Q H\Big(\bigcup_{j=0}^{n} (\mathbf{z}_j, \mathbf{z}_{j+1})\Big)\Big) \\ &> \exp(-C_Q H(A)). \end{aligned}$$

So (6.4) follows from (5.3). □

Acknowledgment

We thank the referees for their input, and one referee in particular for alerting us to the existence of an improvement of Theorem 2.2 over our first version.

References

I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab. **3** (1975), 146–158.

J.-D. Deuschel and D. Stroock, *Large Deviations*, Academic Press, Boston, MA, 1989. Pure & Appl. Math. 137.

J.-D. Deuschel and O. Zeitouni, *Limiting curves for IID records*, Report, 1994, Dept. Math., Technical Univ., Straße des 17 Juni 135, Berlin, Germany; Adv. Appl. Probab. (to appear).

W. Feller, *An Introduction to Probability and its Applications*, vol. I, 3rd edn., revised printing, John Wiley & Sons, New York, 1970.

A. V. Gnedin, *Conical extremes of a multivariate sample*, Report, 1993, Universität Göttingen, Institut für Mathematische Stochastik, Lotzestrasse 13, 3400 Göttingen, Germany; Proc. Conf. Extreme-value Theory, Gaithersburg, May 1993 (to appear).

A. V. Gnedin, *On a best-choice problem with dependent criteria*, J. Appl. Probab. **31** (1994a), 221–234.

A. V. Gnedin, *Multiple maxima of a normal sample*, Report, 1994b, Universität Göttingen, Institut für Mathematische Stochastik, Lotzestrasse 13, 3400 Göttingen, Germany.

C. M. Goldie and R. G. E. Pinch, *Communication Theory*, Cambridge Univ. Press, Cambridge, UK, 1991. London Math. Soc. Student Texts 20.

C. M. Goldie and S. I. Resnick, *Records in a partially ordered set*, Ann. Probab. **17** (1989), 678–689.

C. M. Goldie and S. I. Resnick, *Ordered independent scattering*, Report, 1995, School of Math. Sci., Queen Mary & Westfield College, Mile End Road, London E1 4NS, UK.

A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979. Math. in Sci. and Engineering 143.

R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970. Princeton Math. Ser.

W. Whitt, *Some useful functions for functional limit theorems*, Math. of O. R. **5** (1980), 67–85.

C. M. Goldie, School of Mathematical Sciences, Queen Mary & Westfield College, Mile End Road, London E1 4NS, UK

S. I. Resnick, Cornell University, ORIE, ETC Building, Ithaca, NY 14853–3801, USA