Ergod. Th. & Dynam. Sys. (2002), **22**, 257–280
DOI: 10.1017/S0143385702000123
© *2002 Cambridge University Press*
Printed in the United Kingdom

How many future measures can there be?

KATALIN MARTON† and PAUL C. SHIELDS‡

† *Mathematics Institute, Hungarian Academy of Sciences*
‡ *Mathematics Department, University of Toledo, Toledo, OH 43606, USA*
(e-mail: pshields@math.utoledo.edu)

(*Received* 12 *April* 1999 *and accepted in revised form* 10 *December* 2000)

Abstract. The question addressed in this paper is the worst-case growth rate, for ergodic processes, in the number of conditional measures on n-steps in the future, given the past, that are a fixed distance apart. It is shown that if error is measured using the variational (i.e. distributional) distance then doubly exponential growth is possible, while if error is measured using the $\bar{d}$-metric then more than exponential growth is possible. The question of whether doubly exponential growth is possible in the $\bar{d}$ case is left open.

1. *Introduction*

Let μ be a stationary, ergodic measure on the space $\{0,1\}^Z$ of doubly infinite binary sequences. For $k \leq n$, let x_k^n denote the sequence $x_k, \ldots, x_n$. For $n > 0$, let $\mu_n(\cdot|x_{-\infty}^0)$ denote the conditional measure on n-steps into the future, given an infinite past $x_{-\infty}^0 = (x_0, x_{-1}, \ldots)$, that is, the probability measure on $\{0,1\}^n$ defined by

$$\mu_n(x_1^n|x_{-\infty}^0) = \lim_{m\to\infty} \frac{\mu(x_{-m}^n)}{\mu(x_{-m}^0)}, \quad x_1^n \in \{0,1\}^n,$$

a limit that exists almost surely by the martingale theorem.

If the process is i.i.d. there is only one future distribution, while if the process is Markov or a function of a Markov chain there are only a finite number of future distributions. This paper is concerned with how far a process can be from these two simple cases as measured by the growth rate in the number of future distributions that are a fixed distance apart.

A set in a metric space (M, D) is ϵ-separated if the distance between any two of its members is at least ϵ. The maximum cardinality of an ϵ-separated set will be denoted by $C(M, D, \epsilon)$. The metric spaces of interest in this paper consist of probability measures on $\{0,1\}^n$ defined by conditioning on the past. Such a set of measures is determined by a set $F \subseteq \{0,1\}^Z$, measurable with respect to the past, namely,

$$M_n(F) = \{\mu_n(\cdot|x_{-\infty}^0) : x_{-\infty}^0 \in F\}.$$

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

Two metrics will be of interest: the variational (or distributional) distance and the $\bar{d}$-distance. The variational distance is defined for probability measures ν and η on $\{0, 1\}^n$ by

$$D_n(\nu, \eta) = |\nu - \eta| = \tfrac{1}{2} \sum_{a_1^n} |\nu(a_1^n) - \eta(a_1^n)|,$$

and the $\bar{d}$-distance is defined by

$$\bar{d}_n(\nu, \eta) = \min_{\lambda} E_\lambda d_n(x_1^n, y_1^n),$$

where the minimum is over all joint distributions λ on $\{0, 1\}^n \times \{0, 1\}^n$ that have ν and η as marginals, where E_λ denotes the expected value with respect to λ, and where $d_n(x_1^n, y_1^n)$ denotes the fraction of indices $i \in [1, n]$ for which $x_i \neq y_i$.

We are only interested in growth rates after the removal of sets of small measure, so we define

$$C_n(D_n, \epsilon) = \min C(M_n(F), D_n, \epsilon),$$

where the minimum is over all past measurable sets F of measure at least $1 - \epsilon$. The definition of $C_n(\bar{d}_n, \epsilon)$ is obtained by replacing D_n by $\bar{d}_n$.

Our first result is that doubly exponential growth in the number of future measures that are a fixed variational distance apart is possible.

THEOREM 1. *There exists an ergodic μ and $\epsilon > 0$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log\log C_n(D_n, \epsilon) > 0. \tag{1}$$

For the $\bar{d}$-case we were not able to obtain double exponentiality, but we can show that more than exponential growth is possible.

THEOREM 2. *There exists an ergodic μ and $\epsilon > 0$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log C_n(\bar{d}_n, \epsilon) = \infty. \tag{2}$$

It would be nice to know whether Theorem 1 holds with D_n replaced by $\bar{d}_n$, for if this were true then, as Steve Kalikow pointed out to us, this would provide an example of a process without the weak Pinsker property. (An ergodic process has the weak Pinsker property if given $\epsilon > 0$ it is a stationary coding of the direct product of an i.i.d. process and a process of entropy at most ϵ, see [**7**].)

The measures μ satisfying (1) and (2), respectively, are constructed by a sharpening of the method used in [**5**] and [**3**]. The details are given in the next two sections, which treat, respectively, the variational metric and $\bar{d}$-metric cases.

Remark 1. Examples where the asymptotic growth rates of $C_n(D_n, \epsilon)$ and $C_n(\bar{d}_n, \epsilon)$ are simple to determine include the following.

(1) For i.i.d. processes, $C_n(D_n, \epsilon) = 1$, since there is only one future measure.

(2) For Markov chains and functions of Markov chains, $C_n(D_n, \epsilon) = O(1)$, since there is only a finite number of future measures.

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

(3) For very weak Bernoulli processes, $C_n(\bar{d}_n, \epsilon) \to 1$ for any $\epsilon > 0$, by the definition of very weak Bernoulli.

(4) For entropy zero processes, $(1/n)\log C_n(D_n, \epsilon) \to 0$ for any $\epsilon > 0$, since the infinite past determines the future and the number of future names is subexponential.

(5) For the K-automorphisms constructed in [**4**], and for Kalikow's T, T^{-1} process, [**2**], $(1/n)\log C_n(\bar{d}_n, \epsilon) \to 0$.

For the processes in [**4, 2**], the stronger variational distance form can, in fact, be shown to hold, namely

$$(1/n)\log C_n(D_n, \epsilon) \to 0.$$

For the T, T^{-1} process, this follows from the fact that the past determines the scenery and, with high probability, walking n steps in the future visits at most $n^{2/3}$ sites. For the K-automorphisms of [**4**] only an asymptotically negligible amount of information is needed to specify the start of an n-block and the conditional distribution on $(n-1)$-blocks is uniform, independent of the past.

In the remainder of this paper, log and exp denote the base-2 logarithm and exponential functions, respectively, sequences will be assumed to be binary, unless stated otherwise, and product notation, $uv = (u_1, \ldots, u_n, v_1, \ldots, v_m)$, will be used for the concatenation of $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_m)$, a notation extended to sets of finite sequences by the formulas

$$\mathcal{UV} = \{uv : u \in \mathcal{U}, v \in \mathcal{V}\}, \quad \mathcal{U}^n = \mathcal{U}^{n-1}\mathcal{U}, n > 1.$$

2. *The variational construction*

The variational construction uses the idea that an ergodic process can be inductively defined by describing how finite sequences are to be concatenated to form longer finite sequences, producing in the limit the infinite sample paths. For each ℓ, the joint distribution of ℓ-blocks is obtained by determining the empirical distribution of ℓ-blocks in the sequences of a given stage, then passing to the limit as the stage increases. Ergodicity is guaranteed by making sure that the sequences at any stage occur with almost constant frequency in most members of the next stage. Our particular construction alternates two steps at each stage of the construction, a separation step and a merging step. The separation step creates many different distributions by selecting many different subsets and concatenating separately within each. The resulting sets are merged by forming their concatenation product.

In outline, the separation step starts with a set $\mathcal{S}$ of binary sequences of length k and cardinality n. Subsets of $\mathcal{S}$ of exponentially slightly smaller cardinality are selected then extended by independently concatenating within each subset roughly n/k times. For an appropriate choice of parameters the number of such extended sets will be doubly exponential in k. Even more important is that doubly exponentially many of the original subsets will have small pairwise overlap; this, in turn, guarantees that their extensions give far-apart future measures.

Merging is accomplished by taking the concatenation product $\mathcal{S}'$ of the extended sets in some order. This guarantees that each sequence in $\mathcal{S}$ appears with relative frequency (roughly) $|\mathcal{S}|^{-1}$ in most members of $\mathcal{S}'$, so that iteration of the separation and merging steps yields an ergodic process in the limit.

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

Details are given in the following subsections.

2.1. *Separation and merging.* The input to the separation step is a set $\mathcal{S}$ of k-length binary sequences of cardinality n, together with a positive number $\delta = C(\log n)^{-1/2}$ with the constant C specified later. The relation between k, n, and δ will be specified later; for now the reader should think of n as large and $k = O(\log n)$. For convenience in later arguments it is also assumed that each member of $\mathcal{S}$ starts with a '1'.

A *heap*, more precisely, an $(\mathcal{S}, \delta)$ *heap*, is a subset $\mathcal{H}$ of $\mathcal{S}$ of cardinality $b = \lfloor n^{1-\delta} \rfloor$. A key fact is that m, the number of heaps, is more than exponential in $n^{1-\delta}$; indeed, by using the standard bound (A.3) of Appendix A,

$$m = \binom{n}{\lfloor n^{1-\delta} \rfloor} \tag{3}$$

is asymptotically lower bounded by $2^{cn^{1-\delta}\sqrt{\log n}}$ in the case $\delta = C(\log n)^{-1/2}$, where c is a positive constant. The related formula for the number $\eta(\mathcal{S})$ of heaps that contain a given sequence in $\mathcal{S}$,

$$\eta(\mathcal{S}) = \binom{n-1}{\lfloor n^{1-\delta} \rfloor - 1}, \tag{4}$$

will be useful later.

The *marker* or *header* of a heap $\mathcal{H}$ is the concatenation

$$h(\mathcal{H}) = 10^k w(1)w(2)\cdots w(b),$$

that is, the symbol 1 followed by k consecutive zeros, followed by the members $\{w(i)\}$ of $\mathcal{H}$ in some order. The *extension* $\mathcal{H}^*$ of the heap $\mathcal{H}$ is the set of sequences that can be expressed as a concatenation of the form

$$h(\mathcal{H})v(1)v(2)\cdots v(t), \quad v(i) \in \mathcal{H},$$

where t is the smallest integer such that the total length is at least n. Thus $\mathcal{H}^*$ consists of the sequences that start with the marker after which comes an arbitrary member of the concatenation $\mathcal{H}^t$ of t copies of $\mathcal{H}$.

The output of the separation step is the collection of extensions of the $(\mathcal{S}, \delta)$ heaps, that is,

$$\mathcal{M} = \{\mathcal{H}_1^*, \mathcal{H}_2^*, \ldots, \mathcal{H}_m^*\},$$

ordered in some way. The *merging of* $\mathcal{M}$ is the concatenation product

$$\mathcal{S}' = \mathcal{H}_1^* \cdot \mathcal{H}_2^* \cdot \cdots \cdot \mathcal{H}_m^*, \tag{5}$$

that is, the set of all concatenations of the form

$$u(1)u(2)\cdots u(m), \quad u(r) \in \mathcal{H}_r^*,\ 1 \le r \le m.$$

Note that the members of $\mathcal{S}'$ begin with a '1', so that, in particular, the construction can be repeated with $\mathcal{S}$ replaced by $\mathcal{S}'$.

Recall that k and n are the sequence length and number of sequences, respectively, at the beginning of the separation step. Let b be the heap size, let t be the number of

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

concatenations of a heap used to form an extended heap, let h be the length of the sequences in an extended heap, let m be the number of heaps, and let K and N be the length of the sequences and number of sequences, respectively, after the merging step. These parameters are related as follows.

$$\begin{aligned} &b = \lfloor n^{1-\delta} \rfloor, \quad \delta = C(\log n)^{-1/2}, \\ &h = tk + kb + k + 1, \quad n \le h < n + k, \\ &m \ge \exp(n^{1-\delta}), \\ &K = mh, \\ &N = (b^t)^m. \end{aligned} \tag{6}$$

2.2. *The inductive construction.* The inductive construction begins with a collection $\mathcal{S}(1) \subseteq \{0,1\}^{k(1)}$ of cardinality $n(1)$, each of whose members starts with a '1', and a decreasing sequence $\{\delta(j) : j = 1, 2, \dots\}$ of positive numbers. Repeated alternation of separation and merging produces the *alternating separation–merging sequence*

$$\mathcal{S}(1), \mathcal{M}(1), \mathcal{S}(2), \mathcal{M}(2), \dots, \mathcal{S}(j), \mathcal{M}(j), \dots, \tag{7}$$

where $\mathcal{M}(j)$ consists of the extensions of the $(\mathcal{S}(j), \delta(j))$ heaps and $\mathcal{S}(j+1)$ is the merging of $\mathcal{M}(j)$, for $j = 1, 2, \dots$. The cardinality of $\mathcal{S}(j)$ and the length of its members are denoted by $n(j)$ and $k(j)$, respectively, the cardinality of $\mathcal{M}(j)$ is denoted by $m(j)$, and the length of the sequences that belong to the heap extensions in $\mathcal{M}(j)$ is denoted by $h(j)$. Note, in particular, that $k(j+1) \sim n(j)m(j)$.

The alternating separation–merging sequence (7) defines an ergodic measure μ on $\{0,1\}^Z$, provided only that $\delta(j) \to 0$. This is the measure defined as follows. First, for each ℓ and $k(j) \ge \ell$, define the empirical distribution of overlapping ℓ-blocks in $x_1^{k(j)} \in \mathcal{S}(j)$ by the formula

$$\hat{p}_\ell(a_1^\ell | x_1^{k(j)}) = \frac{|\{i \in [1, k(j) - \ell] : x_{i+1}^{i+\ell} = a_1^\ell\}|}{k(j) - \ell + 1}, \quad a_1^\ell \in A^\ell,$$

and set

$$\hat{p}_\ell(a_1^\ell | \mathcal{S}(j)) = \frac{1}{n(j)} \sum_{x_1^{k(j)} \in \mathcal{S}(j)} \hat{p}_\ell(a_1^\ell | x_1^{k(j)}), \quad a_1^\ell \in A^\ell.$$

The joint distribution of ℓ-blocks is defined as

$$\mu_\ell(a_1^\ell) = \lim_{j \to \infty} \hat{p}_j(a_1^\ell | \mathcal{S}(j)), \quad a_1^\ell \in A^\ell, \tag{8}$$

a limit whose existence will be discussed in a moment. The usual Kolmogorov construction produces a stationary measure μ for which $\mu(\{x : x_1^\ell = a_1^\ell\}) = \mu_\ell(a_1^\ell)$, for each ℓ and a_1^ℓ. The measure μ will be called *the distribution defined by the alternating separation–merging sequence* (7).

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

2.3. *Joint distributions and ergodicity.* It can be shown via direct sequence arguments that the limit (8) indeed exists and that the measure μ it defines is ergodic. We instead make use of an equivalent geometric method called cutting and stacking as it is useful in both the variational and $\bar{d}$ constructions and provides a framework that might be useful for more general constructions. We make free use of the cutting and stacking terminology and theory presented in [**6**, §I.10].

The set $\mathcal{S}(j)$ of $k(j)$-sequences can be thought of as a column structure (also known as a gadget) with columns of uniform width labeled by the members of $\mathcal{S}(j)$, where the column levels are a disjoint collection of left-closed, right-open subintervals of the unit interval $[0, 1)$. In particular, each level of each column has a label, called its name. Associated with this column structure is the upward mapping $T = T_{\mathcal{S}(j)}$ defined on the union of all the levels below the top by the requirement that T map points directly upwards one level.

Let $\eta(\mathcal{S}(j))$ be the number of heaps that contain a given sequence in $\mathcal{S}(j)$, a number not depending on the sequence, see formula (4). Separation is performed by cutting each column of $\mathcal{S}(j)$ into $\eta(\mathcal{S}(j))$ subcolumns of equal width and reassembling to form the heaps $\mathcal{H}_1, \ldots, \mathcal{H}_{m(j)}$, where a heap is obtained by selecting one subcolumn from each of $\lfloor n(j)^{1-\delta(j)} \rfloor$ of the original columns. The extension $\mathcal{H}_r^*$ is formed by $t(j)$-fold independent cutting and stacking of $\mathcal{H}_r$, then stacking the result onto a single column, which is labeled by the heap marker and called its marker column. The number $t(j)$ is chosen to be the smallest integer for which the height $h(j)$ of $\mathcal{H}_r^*$ is at least $n(j)$.

The merging step forms the next stage column structure $\mathcal{S}(j+1)$ by successive independent cutting and stacking of the members of $\mathcal{M}(j)$. Note that the upward map $T_{\mathcal{S}(j+1)}$ is an extension of the upward map $T_{\mathcal{S}(j)}$.

The length of the heap marker added at each stage is negligible relative to the length of the heap extension, indeed, the total mass added for markers at stage $j+1$ can be shown to be summable in j, see (16). The total mass of the top level of each $(j+1)$-stage column is also summable in j, and hence the common extension of all the upward maps defines a transformation T on the unit interval that preserves Lebesgue measure; in particular, the limit (8) exists.

The next step is to show that the measure μ defined by (8) is ergodic. In a slight abuse of notation, the same symbol will be used to denote both a column (or column structure) and its support, that is, the union of its levels (respectively, the union of the levels of the columns that belong to it). The context will usually make clear how the notation should be interpreted. The Lebesgue measure will be denoted by λ, so that, for example, $\lambda(\mathcal{C}|\mathcal{D})$ denotes the Lebesgue measure of the intersection of the supports of $\mathcal{C}$ and $\mathcal{D}$, divided by the Lebesgue measure of the support of $\mathcal{D}$. To establish ergodicity, it is enough to show that structures at later stages become almost independent of earlier stages, that is, it is enough to prove the following lemma.

LEMMA 1.

$$\lim_{j\to\infty} \sum_{\mathcal{C}\in\mathcal{S}(j)} \sum_{\mathcal{D}\in\mathcal{S}(j+1)} |\lambda(\mathcal{C}\cap\mathcal{D}) - \lambda(\mathcal{C})\lambda(\mathcal{D})| = 0.$$

Proof. Fix j, put $\mathcal{S} = \mathcal{S}(j)$, $\delta = \delta(j)$, $n = |\mathcal{S}|$, $t = t(j)$, and $m = m(j)$, and let $\{\mathcal{H}_r : r \le m\}$ be the set of $(\mathcal{S}, \delta)$ heaps. Fix $\mathcal{C} \in \mathcal{S}$, and for $\mathcal{D} \in \mathcal{S}(j+1)$, let $N(\mathcal{C}|\mathcal{D})$

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

denote the number of times a subcolumn of $\mathcal{C}$ was selected in forming $\mathcal{D}$. In particular, if D^* denotes the union of the levels of $\mathcal{D}$ that lie above its marker column, then

$$N(\mathcal{C}|\mathcal{D}) = mt\lambda(\mathcal{C}|\mathcal{D}^*). \tag{9}$$

Relative to the experiment of picking a column $\mathcal{D} \in \mathcal{S}(j+1)$ at random according to the uniform distribution, the random variable $N(\mathcal{C}|\mathcal{D})$ has a binomial (n, p) distribution with $n = \eta(\mathcal{S})t$, $p = \lfloor n^{1-\delta} \rfloor^{-1}$, where $\eta(\mathcal{S})$ denotes the number of heaps that contain a copy of $\mathcal{C}$, hence we have the expected value and variance formulas,

$$E(N(\mathcal{C}|\mathcal{D})) = \frac{\eta(\mathcal{S})t}{\lfloor n^{1-\delta} \rfloor}, \quad V(N(\mathcal{C}|\mathcal{D})) = \frac{\eta(\mathcal{S})t(\lfloor n^{1-\delta} \rfloor - 1)}{\lfloor n^{1-\delta} \rfloor^2},$$

so the counting formula (9) yields

$$E(\lambda(\mathcal{C}|\mathcal{D}^*)) = \frac{\eta(\mathcal{S})t}{mt\lfloor n^{1-\delta} \rfloor}, \quad V(\lambda(\mathcal{C}|\mathcal{D}^*)) = \frac{\eta(\mathcal{S})t(\lfloor n^{1-\delta} \rfloor - 1)}{m^2t^2\lfloor n^{1-\delta} \rfloor^2}. \tag{10}$$

However, $E(\lambda(\mathcal{C}|\mathcal{D}^*)) = \lambda(\mathcal{C}|\mathcal{S}(j+1)^*)$, where $\mathcal{S}(j+1)^*$ is the union of the $\mathcal{D}^*$ for $\mathcal{D} \in \mathcal{S}(j+1)$, so the Cauchy–Schwarz inequality gives the bound

$$E(|\lambda(\mathcal{C}|\mathcal{D}^*) - \lambda(\mathcal{C}|\mathcal{S}(j+1)^*)|) \le \sqrt{V(\lambda(\mathcal{C}|\mathcal{D}^*))}.$$

Substitution of the mean and variance formulas (10) and simple algebra using the formulas (3) and (4) for m and $\eta(\mathcal{S})$, respectively, then yield

$$nE(|\lambda(\mathcal{C}|\mathcal{D}^*) - \lambda(\mathcal{C}|\mathcal{S}(j+1)^*)|) \to 0, \quad \text{as } j \to \infty.$$

The left-hand side of this is, after ignoring the negligible contribution of the markers, the same as

$$\sum_{\mathcal{C}\in\mathcal{S}(j)} \sum_{\mathcal{D}\in\mathcal{S}(j+1)} |\lambda(\mathcal{C} \cap \mathcal{D}) - \lambda(\mathcal{C})\lambda(\mathcal{D})|,$$

which therefore must go to 0, as $j \to \infty$. This completes the proof of Lemma 1, that is, μ is indeed ergodic. □

2.4. *Future distributions.* Our first task is to show that when looking at the same level of the heap extensions formed from heaps with small overlap, future measures, conditioned on the past, must be far apart in variational distance. Towards this end, we first make a few observations about points in the unit interval and their names.

For almost every x in the unit interval there is, for every integer n, a stage $j = j(x, n)$ such that $T^n x$ belongs to the support of $\mathcal{S}(j)$. The name of such an x is the sequence $\{x_n\}$ where x_n is the label of the level of the column of $\mathcal{S}(j(x, n))$ to which $T^n x$ belongs. The mapping $x \mapsto \{x_n\}$ is an invertible (i.e. almost surely one-to-one and onto) measurable mapping of the unit interval into the space of doubly-infinite binary sequences that carries Lebesgue measure λ onto μ and T onto the shift transformation.

Invertibility of $x \mapsto \{x_n\}$ follows from the assumption that the initial set $\mathcal{S}(1)$ of $k(1)$-length sequences all start with a 1; the construction then guarantees that no block of the form 10^k1, with $k > k(1)$ can occur in a name unless $k = k(j)$ for some j and it is

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

the beginning of the marker associated with an extended heap $\mathcal{H}_r^*$ formed from $\mathcal{S}(j)$. Note also that a block $10^{k(j)}1$ cannot start twice in any time interval of length $h(j)$. If $x \neq y$ then there is a j such that they either belong to different levels of the heap extensions formed from $\mathcal{S}(j)$, in which case the block $10^{k(j)}1$ starts in different places in the first $h(j)$ places of the past of each, or to the same level of different heap extensions, in which case, the heap markers disagree yet start at the same place in the first $h(j)$ places of the past of each.

In a slight abuse of notation, justified in part by the invertibility of the mapping $x \mapsto \{x_n\}$, the same symbol, such as x, will be used for a point in the unit interval and its name $\{x_n\}$. Likewise, the same symbol, such as B, will be used for a subset of the sequence space $\{0, 1\}^Z$ and for the set of points in the unit interval whose names belong to B.

Fix a stage j and let $k = k(j)$, $n = n(j)$, $m = m(j)$, $h = h(j)$, and $\delta = \delta(j)$. For each L in the range

$$k\lfloor n^{1-\delta}\rfloor + k + 1 < L \leq h - 2k, \tag{11}$$

let $[L]$ denote the union of the intervals that lie at level L of the columns of the heap extensions $\{\mathcal{H}_i^*\}$ of the $(\mathcal{S}, \delta)$ heaps. Property (11) ensures that level L is above the marker levels and below the top $2k$ levels. Define $r(x) = r$ if x lies in the support of $\mathcal{H}_r^*$, and note that the value of r is completely determined by the past x_{-L+1}^0.

LEMMA 2. *If $x, y \in [L]$ and $|\mathcal{H}_{r(x)} \cap \mathcal{H}_{r(y)}| \leq \delta|\mathcal{H}_{r(x)}|$ then*

$$|\mu_{2k}(\cdot|x_{-\infty}^0) - \mu_{2k}(\cdot|y_{-\infty}^0)| \geq 1 - \delta.$$

Proof. Let s be the least non-negative integer such that $L + s - (k\lfloor n^{1-\delta}\rfloor + k + 1)$ is a multiple of k. Following the marker of length $k\lfloor n^{1-\delta}\rfloor + k + 1$, k-blocks from the heap $\mathcal{H}_i$ are concatenated to form the heap extension $\mathcal{H}_i^*$, in particular the definition of s guarantees that the distribution of k-blocks that start at $L + s + 1$ is uniform over the heap $\mathcal{H}_i$. In other words, letting μ_{2k}^s denote the marginal of a measure μ_{2k} on A^{2k} onto its coordinates from $s + 1$ to $s + k$, that is, the measure on k-blocks defined by

$$\mu_{2k}^s(a_{s+1}^{s+k}) = \sum_{a_1^s, a_{s+k+1}^{2k}} \mu_{2k}(a_1^{2k}),$$

we have

$$\mu_{2k}^s(a_{s+1}^{s+k}|x_{-\infty}^0) = \begin{cases} |\mathcal{H}_{r(x)}|^{-1}, & \text{if } a_{s+1}^{s+k} \in \mathcal{H}_{r(x)}, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

The lemma now follows from the contraction property of variational distance, i.e. $|\mu_{2k} - \nu_{2k}| \geq |\mu_{2k}^s - \nu_{2k}^s|$. □

2.5. *Doubly-exponential growth.* It remains to show that for an appropriate choice of $k(1)$, $n(1)$, and the constant in the definition of $\delta(j)$, the measure μ satisfies the doubly-exponential growth condition (1). The key to this is a combinatorial fact about heap constructions, formulated here in terms of abstract sets. Fix a set $\mathcal{S}$ of cardinality n and $0 < \delta < 1$. In this abstract setting a heap is just a subset $\mathcal{H}$ of $\mathcal{S}$ of cardinality $\lfloor n^{1-\delta}\rfloor$. A collection of heaps is said to be $(1-\delta)$-*pairwise-separated*, if the intersection of any two members of the collection has cardinality at most $\delta\lfloor n^{1-\delta}\rfloor$. For each n, let $C_n^*(\delta)$ denote

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

the maximum cardinality of a $(1-\delta)$-pairwise-separated collection of heaps. The proof of the following lemma is given in Appendix A.

LEMMA 3. *There are constants C and n_0 such that if $n \geq n_0$ and $\delta = C(\log n)^{-1/2}$ then*

$$C_n^*(\delta) \geq \exp(n^{1-\delta}). \tag{13}$$

We are now ready to establish our desired variational result, stated here as the following proposition.

PROPOSITION 1. *Given $0 < H < 1$, there is an integer $k(1)$, a positive constant C, and a set $\mathcal{S}(1)$ of $k(1)$-length binary sequences of cardinality $n(1) \geq 2^{k(1)H}$, such that each member of $\mathcal{S}(1)$ starts with a 1 and such that if $\delta(j) = C(\log n(j))^{-1/2}$, for $j \geq 1$, then the alternating separation–merging sequence (7) defines an ergodic measure μ for which the doubly-exponential growth property (1) holds.*

Proof. Fix $0 < H < 1$ and let n_0 and C be the numbers given by Lemma 3. Choose k_0 so large that if $k(1) \geq k_0$ there is a set $\mathcal{S}(1) \subset \{0, 1\}^{k(1)}$ of cardinality at least $2^{k(1)H}$ such that each member of $\mathcal{S}(1)$ starts with a '1'. Given such a $k(1)$ and $\mathcal{S}(1)$, let $\{\mathcal{S}(j)\}$ be the alternating separation–merging sequence using $\delta(j) = C(\log n(j))^{-1/2}$, $j \geq 1$, and let μ be the measure defined by this sequence. The goal is to show that $k(1)$ can be chosen so large that the doubly-exponential growth property (1) holds. As a step towards this end we first establish some useful facts about the parameters, stated as the following lemma.

LEMMA 4. *If $k(1)$ is large enough then:*

(i) $k(j) \leq (\log n(j))^2$, $j \geq 1$;

(ii) $\dfrac{\log n(j+1)}{k(j+1)} \geq (1-\delta(j))^2 \dfrac{\log n(j)}{k(j)}$, $j \geq 1$; *and*

(iii) $\displaystyle\sum_j \delta(j) < \infty$.

Proof. The proof of (i) is by induction. To get started note that $k(1) = (1/H)\log n(1)$ is certainly dominated by $(\log n(1))^2$, if $k(1)$ is large enough, since H is fixed. Assume $k(j) \leq (\log n(j))^2$. The formulas (6) relate the parameters at stage $j+1$ with those at stage j. In particular, $K = mh$ and $h \leq n + k$ yield

$$k(j+1) = m(j)h(j) \leq 2m(j)n(j).$$

If $k(1)$ is large enough then $\delta(1) \leq 1/2$, and hence $m \geq \exp(n^{1-\delta})$ yields $m \geq \exp(n^{1/2})$. Since $t(j)\log b(j) \geq t(1)\log b(1)$ and we can assume the latter is at least 1, it follows that

$$2n(j) \leq m(j)(t(j)\log b(j))^2$$

will hold if $k(1)$ is large enough, in which case,

$$k(j+1) \leq 2m(j)n(j) \leq m(j)^2(t(j)\log b(j))^2 = (\log n(j+1))^2,$$

since $n(j+1) = b(j)^{m(j)t(j)}$ by (6). This proves that (i) holds if $k(1)$ is large enough.

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

To establish (ii) we use the parameter formulas (6) to express it in the form

$$\frac{\log N}{K} = \frac{tk \log b}{h \log n^{1-\delta}} \geq (1-\delta)$$

which, after some algebra, takes the form

$$\frac{h-tk}{h} + \frac{tk}{h \log n^{1-\delta}} \log \frac{n^{1-\delta}}{b} \leq \delta. \tag{14}$$

Thus it is enough to show that each of these terms is $o(\delta)$.

Since $h \geq n$ and $b \leq n^{1-\delta}$ we have

$$\frac{h-tk}{h} = \frac{kb+k+1}{h} \leq \frac{kn^{1-\delta}+2k+1}{n} \leq 4kn^{-\delta} \leq 4n^{-\delta}\log^2 n,$$

by property (i). This final bound is certainly $o(\delta)$ since $n^{-\delta} = 2^{-\sqrt{\log n}}$.

For the second term of (14) we use $n^{1-\delta} \leq b+1$ and $tk \leq h$ to obtain

$$\frac{tk}{h \log n^{1-\delta}} \leq \frac{\log(1+1/b)}{\log n^{1-\delta}} \leq \frac{1/(b \ln 2)}{\log n^{1-\delta}} \leq \frac{1/\ln 2}{\sqrt{n}\log\sqrt{n}}$$

once $k(1)$ is large enough to guarantee $1-\delta(j) \geq 1/2$ for all j. Since this final bound is $o(\delta)$ this shows that (14) holds for $k(1)$ large enough and completes the proof of property (ii).

To establish (iii), we first use part (ii) to obtain

$$\log n(j+1) \geq (k(j+1)/k(j))(1-\delta(j))^2 \log n(j),$$

then use the fact that $k(j+1) = m(j)h(j) \geq m(j)n(j)$, that $n(j)/k(j) \geq 1$, that $\delta(j) \geq 1/2$, and that $m(j) \geq \exp(n(j)^{1/2})$, to obtain

$$\log n(j+1) \geq (1/4)\exp(n(j)^{1/2}) \log n(j),$$

in other words,

$$\frac{1}{\sqrt{\log n(j+1)}} \leq 4\exp(n(j)^{-1/2}) \frac{1}{\sqrt{\log n(j)}}.$$

This shows that $\delta(j+1) \leq (1/2)\delta(j)$, if $n(j)$ is large enough, hence that (iii) holds if $k(1)$ is large enough. □

We assume that $k(1)$ is large enough so the properties of Lemma 4 hold. Our goal is to show the existence of many far apart future measures even after a past set of small measure has been removed. The precise statement is as follows.

LEMMA 5. *Given a past measurable subset $B \subset \{0,1\}^Z$ of measure $\gamma < 1/2$, for all sufficiently large j there is a finite set G, disjoint from B, such that*

(i) $|G| \geq (1-2\gamma)\exp(n(j)^{1-\delta(j)})$*; and*

(ii) $|\mu_{2k(j)}(\cdot|x_{-\infty}^0) - \mu_{2k(j)}(\cdot|y_{-\infty}^0)| \geq 1-\delta$*, if* $x, y \in G$, $x \neq y$.

Proof. First consider a stage j. Let $k = k(j)$, $n = n(j)$, $\delta = \delta(j)$, $b = b(j)$, $t = t(j)$, $h = h(j)$, $m = m(j)$, and $\mathcal{S} = \mathcal{S}(j)$, and let $\mathcal{M} = \mathcal{M}(j) = \{\mathcal{H}_1^*, \mathcal{H}_2^*, \ldots, \mathcal{H}_m^*\}$. For each $I \subseteq \{1, 2, \ldots, m\}$, let $\mathcal{M}_I = \{\mathcal{H}_i^* \in \mathcal{M} : i \in I\}$. Lemma 3 and the choice of C provide

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

an I of cardinality at least $\exp(n^{1-\delta})$ for which $\mathcal{M}_I$ is $(1-\delta)$-pairwise-separated, where two heap extensions are $(1-\delta)$-separated if their underlying heaps are $(1-\delta)$-separated. Fix such an I.

A permutation ϕ of $\mathcal{S}$ extends to the permutation $\hat{\phi}$ of $\mathcal{M}$ defined by $\hat{\phi}(\mathcal{H}_i^*) = (\phi(\mathcal{H}_i))^*$. In particular, the collection

$$\mathcal{M}_\phi = \{\hat{\phi}(\mathcal{H}_i^*) : i \in I\},$$

is also a $(1-\delta)$-pairwise-separated collection of the same cardinality as $\mathcal{M}_I$. Furthermore, if $\Pi(\mathcal{S})$ is the set of all permutations of $\mathcal{S}$, then the collection $\{\mathcal{M}_\phi : \phi \in \Pi(\mathcal{S})\}$ is an even cover of $\mathcal{M}$ in the sense that each $\mathcal{H}_i^*$ is contained in exactly the same number of $\mathcal{M}_\phi$.

Let $B \subset \{0,1\}^Z$ be a past measurable set of measure $\gamma < 1/2$. Let B also denote the set of points in the unit interval whose names belong to B and note that $\lambda(B) = \gamma$. As earlier, for L in the range

$$k\lfloor n^{1-\delta}\rfloor + k + 1 < L \le h - 2k, \tag{15}$$

let $[L]$ denote the union of the intervals that lie at level L of the columns that belong to members of $\mathcal{M}$ and define $r(x) = r$ if x lies in the support of $\mathcal{H}_r^*$. The total mass of the top $2k$ and bottom $k\lfloor n^{1-\delta}\rfloor + k + 1$ levels of the $\mathcal{H}_i^*$, is summable in j, for Lemma 4(i) and $h \le n$ imply that

$$\frac{k\lfloor n^{1-\delta}\rfloor + k + 1 + 2k}{h} \le 5\frac{kb}{h} \le 5\frac{kn^{1-\delta}}{n} \le 5\frac{\log^2 n}{n^\delta} \le 5\delta, \tag{16}$$

if $k(1)$ is large enough, so Lemma 4(iii) gives the desired summability in j.

It follows that if j is large enough then there must be an L in the range (15) for which

$$\lambda(B \cap [L]) \le 2\gamma\lambda([L]).$$

Fix such an L. Since $\{\mathcal{M}_\phi : \phi \in \Pi(\mathcal{S})\}$ is an even cover of $\mathcal{M}$, there must be an $\mathcal{M}_\phi$ for which

$$\lambda(B \cap [L] \cap [\mathcal{M}_\phi]) \le 2\gamma\lambda([L] \cap [\mathcal{M}_\phi]), \tag{17}$$

where $[\mathcal{M}_\phi]$ denotes the union of the supports of the $\mathcal{H}_i^*$ that belong to $\mathcal{M}_\phi$.

Let $\mathcal{M}'_\phi$ be the subset of $\mathcal{M}_\phi$ consisting of those $\mathcal{H}_i^*$ for which $\mathcal{H}_i^* \cap [L]$ has a non-empty intersection with the complement of B. Condition (17) guarantees that $|\mathcal{M}'_\phi| \ge (1-2\gamma)\exp(n^{1-\delta})$. Thus we can select one point from each set $\mathcal{H}_i^* \cap [L] \cap B^c$, $\mathcal{H}_i^* \in \mathcal{M}'_\phi$, to obtain a set G with the following properties.

(a) G is a subset of $[L]$ of cardinality at least $(1-2\gamma)\exp(n^{1-\delta})$ and is disjoint from B.

(b) The collection $\{\mathcal{H}_{r(x)} : x \in G\}$ is $(1-\delta)$-pairwise-separated.

Property (b) follows from the fact that $\{\mathcal{H}_{r(x)}, x \in G\}$ is a subset of the well-separated collection $\mathcal{M}_\phi$; in particular, it implies that

$$|\mu_{2k}(\cdot|x_{-\infty}^0) - \mu_{2k}(\cdot|y_{-\infty}^0)| \ge 1 - \delta, x, y \in G, x \ne y,$$

by Lemma 2. This completes the proof of Lemma 5. □

The lemma implies that even after the removal of a past set of measure less than $1/4$, for all sufficiently large j there will be at least $(1/2)\exp(n(j)^{1-\delta(j)})$ future measures on

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

$2k(j)$-length blocks that are mutually at least $1 - \delta(j)$ apart in variational distance. It remains to be shown that $\exp(n(j)^{1-\delta(j)})$ is doubly exponential in $k(j)$, as $j \to \infty$. Towards this end, note that if

$$\alpha(j) = (1 - \delta(1))(1 - \delta(2)) \cdots (1 - \delta(j-1)),$$

then $\alpha = \lim_j \alpha(j)$ exists and is positive by Lemma 4(iii). Furthermore, part (ii) of that lemma implies that

$$\frac{\log n(j)}{k(j)} \geq \alpha^2 \frac{\log n(1)}{k(1)} = \alpha^2 H, \quad j \geq 1,$$

which implies that

$$\log\log[\exp(n(j)^{1-\delta(j)})] = (1 - \delta(j)) \log n(j) \geq \left(\frac{1 - \delta(j)}{2}\right) \alpha^2 H \cdot 2k(j),$$

that is, $\exp(n(j)^{1-\delta(j)})$ is doubly exponential in $2k(j)$, since $\delta(j) \to 0$. This completes the proof of Proposition 1. □

3. *The $\bar{d}$ construction*

Getting a $\bar{d}$ construction started is easy, for if $\mathcal{S} \subseteq \{0, 1\}^k$ has the property that any two of its members are $\bar{d}$-far apart, then two heaps with small enough overlap produce future measures that are almost as far apart in $\bar{d}$. The problem comes in the merging step, for while the simple merging defined by (5) produces distinct sequences, the sequences it produces need not be $\bar{d}$-far apart, hence repetition of the construction will not produce a large number of far-apart measures at the next stage.

We adopt a trio of strategies to get around the difficulties raised in the preceding paragraph.

(1) We focus on well-separated *sets of sequences* rather than individual sequences. In particular, heaps will be collections of well-separated sets.

(2) We restrict attention to collections of heaps that are pairwise well-separated. This restriction causes some stationarity problems as it is no longer true that each set appears in the same number of heaps, but these can be overcome by using random selection to insure that each set appears in almost the same number of heaps.

(3) Rather than using one merging pattern we use a large number of different patterns, choosing them so that any pair of patterns agree in at most a small fraction of places. This, our most important change, is used to produce a large number of well-separated sets for the next stage of the construction.

3.1. *Collections of well-separated sets.*

Let $\mathcal{S} = \{A_i : i \leq n\}$ be a collection of disjoint subsets of $\{0, 1\}^k$ of equal cardinality. *Note that in this new setting, $\mathcal{S}$ is a collection of subsets, not a collection of sequences.* Such a collection is said to be (ϵ, θ)-*core-separated* if for each i there is a subset $A_i^\circ \subset A_i$ such that the following hold:

(a) $|A_i^\circ| \geq (1 - \theta)|A_i|, i \leq n$;

(b) $d_k(A_i^\circ, A_j^\circ) \geq \epsilon$, if $i \neq j$.

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

Here $d_k(A_i^\circ, A_j^\circ)$ denotes the distance between the sets A_i° and A_j° defined by the per-symbol Hamming metric $d_k(x_1^k, y_1^k) = (1/k)|\{i \in [1,k] : x_i \neq y_i\}|$. The collection $\{A_i^\circ\}$ is called an (ϵ, θ)-*core of* $\mathcal{S} = \{A_i\}$. The concept of core-separation, which emphasizes the concept of core, makes precise the idea of 'well-separated collections of sets'.

As earlier, our construction is done in stages. A stage begins with an (ϵ, θ)-core-separated collection $\mathcal{S} = \{A_i : i \leq n\}$ of $n \leq 2^{k-1}$ disjoint subsets of $\{0,1\}^k$ of equal cardinality. The values of the natural number n and the positive numbers ϵ and θ depend on the stage of iteration, though it will be important later that $k \leq \log^2 n$, that ϵ stays bounded above zero, and that θ goes to zero very rapidly as the stage increases. The assumption $n \leq 2^{k-1}$ is only to provide enough space to put in markers.

In this new setting, we put $\delta = 1/\sqrt{\log n}$ and define an $(\mathcal{S}, \delta)$ heap to be a subset of $\mathcal{S}$ of cardinality $b = \lfloor n^{1-\delta} \rfloor$. Note, however, that since the members of $\mathcal{S}$ are sets of sequences, not individual sequences, *the elements of a heap are also sets of sequences, not individual sequences*.

In the remainder of this section and in the next section we use the following notation.

$$|\mathcal{S}| = n, \quad \delta = \frac{1}{\sqrt{\log n}}, \quad b = \lfloor n^{1-\delta} \rfloor = \text{cardinality of the } (\mathcal{S}, \delta)\text{-heap.}$$

The first step in going to the next stage is to construct a large number of heaps with small pairwise overlap and with the additional property that the number of heaps containing a given A_i is almost constant in i. The latter restriction reduces somewhat the number of heaps we can construct but this will not cause difficulties. The precise result is stated as the next lemma; its proof via random selection is given in Appendix A.

LEMMA 6. *If n is large enough and $m = \lfloor \exp(n^{1-3\delta}) \rfloor$, then there is a collection $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_m\}$ of $(\mathcal{S}, \delta)$ heaps such that*

(a) $|\mathcal{H}_r \cap \mathcal{H}_s| \leq 2n^{1-2\delta}$, $r \neq s$; *and*

(b) $|\,|\{r : A_i \in \mathcal{H}_r\}| - mn^{-\delta}| \leq m^{3/4}\sqrt{n^{-\delta}}$, $i \in \mathcal{S}$.

In the following we assume n is large enough that the lemma holds and that $\{\mathcal{H}_r : 1 \leq r \leq m\}$ denotes a collection of $(\mathcal{S}, \delta)$ heaps which satisfy properties (a) and (b) of the lemma.

For each r, let μ^r denote the uniform distribution on the set $U_r = \cup_{A_i \in \mathcal{H}_r} A_i$ of k-sequences that belong to the members of $\mathcal{H}_r$.

LEMMA 7. *If n is large enough then*

$$\bar{d}_k(\mu^r, \mu^s) \geq \epsilon(1 - 2\theta - 4\delta), \quad r \neq s. \tag{18}$$

Proof. Let $\{A_i^\circ\}$ be a core of $\{A_i\}$. Fix $r \neq s$. The two sets

$$U_r^\circ = \bigcup_{A_i \in \mathcal{H}_r - \mathcal{H}_s} A_i^\circ, \quad U_s^\circ = \bigcup_{A_i \in \mathcal{H}_s - \mathcal{H}_r} A_i^\circ$$

are at least ϵ apart and satisfy

$$\mu^r(U_r^\circ) \geq (1 - \theta - 2n^{-\delta}), \quad \mu^s(U_s^\circ) \geq (1 - \theta - 2n^{-\delta}).$$

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

It follows that any joining of μ^r and μ^s must have at least $(1 - 2\theta - 4n^{-\delta})$ of its mass concentrated on $U_r^\circ \times U_s^\circ$, which, in turn, implies that (18) holds for large enough n, since $n^{-\delta}$ approaches zero faster than $\delta = (\log n)^{-1/2}$. □

From now on we assume n is large enough so that Lemma 6 and Lemma 7 both hold. Our next task is to define heap extensions. Let $t = n^2$. (We take somewhat longer sequences in the $\bar{d}$ case.) In a slight abuse of notation, let $\mathcal{H}_r^t$ denote the set of all binary sequences of the form x_1^{tk}, such that

$$x_{\ell k+1}^{(\ell+1)k} \in \bigcup_{A_i \in \mathcal{H}_r} A_i, \quad \ell = 0, 1, \ldots, t-1.$$

In our variational construction a heap marker was a concatenation of the sequences in the heap with a header that could be unambiguously located. Now, however, $\mathcal{S}$ consists of sets of sequences. Here is where the assumption $|\mathcal{S}| = n \le 2^{k-1}$ is used, for this means we can index the sets in $\mathcal{S}$ by k-length binary sequences, each of which starts with a 1. Fix such an indexing. The *marker* of a heap $\mathcal{H}_r$ is the concatenation

$$h(\mathcal{H}_r) = 10^k w(1)w(2)\cdots w(b),$$

where $b = \lfloor n^{1-\delta} \rfloor = |\mathcal{H}_r|$, and the $w(i)$ are the k-length binary sequences which are the indices of those sets in $\mathcal{S}$ that were selected to form $\mathcal{H}_r$. The members of $\mathcal{H}_r^t$ are extended by prefixing with the marker to form the set

$$\mathcal{H}_r^* = h(\mathcal{H}_r)\mathcal{H}_r^t = \{h(\mathcal{H}_r)x_1^{tk} : x_1^{tk} \in \mathcal{H}_r^t\},$$

which is called the (*marked*) *t-extension of $\mathcal{H}_r$*. Note that the sequences of the marked t-extension all have length $k + 1 + kb + tk$, where $t = n^2$, $b = \lfloor n^{1-\delta} \rfloor$. The collection of (marked) extensions of the $(\mathcal{S}, \delta)$ heaps,

$$\mathcal{M} = \{\mathcal{H}_1^*, \mathcal{H}_2^*, \ldots, \mathcal{H}_m^*\},$$

will be called the *strong heap separation of $\mathcal{S}$*. Note that each $\mathcal{H}_i^*$ is a *set of sequences, not a set of sets of sequences*.

We later need the fact that 'typical' sequences that belong to different heap extensions must be far apart. In making this precise we use concatenation notation for tk-length sequences, writing $x_1^{tk} = \beta_1^t$, where $\beta_\ell = x_{\ell k+1}^{(\ell+1)k}$, $\ell = 0, 1, \ldots, t-1$, and $t = n^2$.

For each r and $\beta_1^t \in \mathcal{H}_r^t$, let $X_{r,i}(\beta_1^t)$ denote the number of indices j such that $\beta_j \in A_i$, for $A_i \in \mathcal{H}_r$, and let $X_r'(\beta_1^t)$ be the number of indices j such that $\beta_j \in \cup_{A_i \in \mathcal{H}_r}(A_i - A_i^\circ)$. These can be regarded as binomial random variables with respective expected values

$$E(X_{r,i}) = \frac{t}{b}, \quad E(X_r') \le \theta t, \quad b = \lfloor n^{1-\delta} \rfloor, \quad t = n^2.$$

A sequence $x_1^{tk} = \beta_1^t \in \mathcal{H}_r^t$ is called *$\mathcal{H}_r$-typical* if the following two conditions hold.

(a) $|X_{r,i}(\beta_1^t) - E(X_{r,i})| \le \delta n^{1+2\delta}$, for all $A_i \in \mathcal{H}_r$.

(b) $|X_r'(\beta_1^t) - E(X_r')| \le \delta n^2$.

The set of typicals has asymptotically large probability, as stated in the following lemma.

LEMMA 8. *If n is large enough, the probability that a randomly chosen $\beta_1^t \in \mathcal{H}_r^t$ is not $\mathcal{H}_r$-typical is at most δ.*

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

Proof. An application of the well-known large deviations bound

$$\text{Prob}(|X - E(X)| \geq \epsilon n) \leq (n+1)\exp(-2n\epsilon^2), \tag{19}$$

valid when X is a binomial (n, p) random variable, yields

$$\text{Prob}(|X_{r,i} - E(X_{r,i})| \geq \delta n^{1+2\delta}) \leq 2n^2 \exp(-2\delta^2 n^{4\delta}),$$

so that

$$\text{Prob}\Big(\max_i |X_{r,i} - E(X_{r,i})| \geq \delta n^{1+2\delta}\Big) \leq 2n^3 \exp(-2\delta^2 n^{4\delta}).$$

Likewise, $P(|X'_r - E(X'_r)| \geq \delta n^2) \leq 2n^2 \exp(-2\delta^2 n^2)$, and hence, if n is large enough, the probability that a randomly chosen $\beta_1^t \in \mathcal{H}_r^t$ is not $\mathcal{H}_r$-typical is bounded above by

$$2n^3 \exp(-2\delta^2 n^{4\delta}) + 2n^2 \exp(-2\delta^2 n^2) \leq \delta,$$

which establishes the lemma. □

The separation property we need is stated as the following lemma.

LEMMA 9. *If n is large enough, if $x_1^{tk} = \beta_1^t$ is $\mathcal{H}_r$-typical, if $y_1^{tk} = \gamma_1^t$ is $\mathcal{H}_s$-typical, and if $r \neq s$, then*

$$d_{tk}(x_1^{tk}, y_1^{tk}) \geq \epsilon(1 - 2\theta - 14\delta).$$

Proof. The definition of (ϵ, Θ)-core-separation implies that $d_k(\beta_i, \gamma_i) \geq \epsilon$, unless the k-blocks β_i and γ_i belong to the same A_j or one of them belongs to

$$\cup_{A_j \in \mathcal{H}_r}(A_j - A_j^\circ), \quad \text{respectively, } \cup_{A_j \in \mathcal{H}_s}(A_j - A_j^\circ).$$

Part (a) of the definition of typicality implies that a k-block lies in a given A_j at most $\delta n^{1+2\delta} + n^2/b$ times, and hence Lemma 6(a) implies that β_i and γ_i can belong to the same A_j for at most

$$2\left(\delta n^{1+2\delta} + \frac{2n^2}{n^{1-\delta}}\right) 2n^{1-2\delta} = (4\delta + 8n^{-\delta})n^2$$

indices i, a quantity bounded above by $12\delta n^2$, if n is large enough. However, typicality implies that one of β_i or γ_i belongs to

$$\cup_{A_j \in \mathcal{H}_r}(A_j - A_j^\circ), \quad \text{respectively, } \cup_{A_j \in \mathcal{H}_s}(A_j - A_j^\circ).$$

for at most $2(E(X'_r) + \delta n^2) \leq 2(\theta + \delta)n^2$ indices. The lemma follows. □

The final step in going from one stage to the next is discussed in the following subsection.

3.2. *Merging via permutations.* The final step in going from one stage to the next is the merging step. In our variational construction, heap extensions were merged in a fixed order. To keep as much separation as possible for the $\bar{d}$ case, we use a large number of merging orderings, subject to the condition that any two orderings are quite different. Two permutations π_1 and π_2 of $\{1, 2, \ldots, m\}$ will be said to be $(1 - \delta)$-*separated* if $\pi_1(\ell) = \pi_2(\ell)$ for at most δm indices $\ell \in \{1, 2, \ldots, m\}$. A proof of the following lemma is given in Appendix A.

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

LEMMA 10. *If n is sufficiently large, $\delta = 1/\sqrt{\log n}$, and $m = \lfloor \exp(n^{1-3\delta}) \rfloor$, there is a collection Π of pairwise $(1-\delta)$-separated permutations of $\{1, 2, \ldots, m\}$ whose cardinality satisfies*

$$|\Pi| \geq \exp\left(\frac{\delta}{2} m \log m\right). \tag{20}$$

Henceforth we assume n is large enough that Lemmas 6–10 all hold. Fix a pairwise $(1-\delta)$-separated collection Π of permutations of $\{1, 2, \ldots, m\}$ such that $|\Pi| = \lceil \exp((\delta/2)m \log m) \rceil$. For each $\pi \in \Pi$, the concatenation product

$$B_\pi = \mathcal{H}^*_{\pi(1)} \cdot \mathcal{H}^*_{\pi(2)} \cdot \cdots \cdot \mathcal{H}^*_{\pi(m)},$$

is called the *π-merging* of the strong heap separation of $\mathcal{S}$. Note that this merging is a subset of $\{0, 1\}^K$, where

$$K = \lfloor \exp(n^{1-3\delta}) \rfloor (k + 1 + k\lfloor n^{1-\delta} \rfloor + n^2 k).$$

The collection $\{B_\pi : \pi \in \Pi\}$ is disjoint since distinct heap extensions have distinct markers; it will be called the *permutation-separated merging of the strong heap extension of $\mathcal{S}$* or, more simply, the *Π-merging of $\mathcal{S}$*.

A key lemma is that the Π-merging $\{B_\pi : \pi \in \Pi\}$ is also core-separated with a smaller ϵ and a suitably controlled θ.

LEMMA 11. *If n is large enough, the Π-merging $\{B_\pi : \pi \in \Pi\}$ of an (ϵ, θ)-core-separated $\mathcal{S}$ is (ϵ', θ')-core-separated, where:*

(i) $\epsilon' \geq \epsilon(1 - 2\theta - 14\delta)(1 - 7\delta)$;

(ii) $\theta' \leq 2m \exp(-2m\delta^2)$.

Proof. Fix $\pi \in \Pi$. For $x_1^K \in B_\pi$ let $\tilde{x}_1^{mtk}$ denote the sequence obtained by deleting the markers of the $\mathcal{H}^*_{\pi(i)}$ and write

$$\tilde{x}_1^{mtk} = u_1 u_2 \cdots u_m, \quad u_i \in \mathcal{H}^t_{\pi(i)}, \quad 1 \leq i \leq m.$$

Let $Y(x_1^K)$ denote the number of indices i such that u_i is $\mathcal{H}_{\pi(i)}$-typical. Lemma 8 and the definition of π-merging imply that Y is a binomial random variable with expected value at least $m(1-\delta)$. Let B^*_π be the set of all $x_1^K \in B_\pi$ for which $|Y(x_1^K) - E(Y)| \leq m\delta$, and define θ' by the formula

$$|B^*_\pi| = (1 - \theta')|B_\pi|.$$

The large deviations bound (19) implies that property (ii) holds, so to complete the proof of Lemma 11 it is enough to show that property (i) also holds provided n is large enough. Suppose $|Y(x_1^K) - E(Y)| \leq m\delta$. The relation $E(Y) \geq m(1-\delta)$ implies

$$Y(x_1^K) \geq m(1-\delta) - m\delta = m(1 - 2\delta). \tag{21}$$

Furthermore, the fraction of x_1^K taken up by the heap markers satisfies

$$\frac{k + 1 + kb}{k + 1 + kb + kt} \leq \frac{3b}{t} \leq \delta, \tag{22}$$

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

if n is large enough, since $b = \lfloor n^{1-\delta} \rfloor$, and $t = n^2$. Thus if $x_1^K \in B_\pi^*$ and $y_1^K \in B_{\pi'}^*$, where $\pi \neq \pi'$, then

$$d_K(x_1^K, y_1^K) \geq \epsilon(1 - 2\theta - 14\delta)(1 - 7\delta), \tag{23}$$

where the first factor comes from Lemma 9, while the second factor comes from the bounds (21) and (22), which must be doubled since they apply to both x_1^K and y_1^K, together with the fact that at most a δ-fraction of heap extensions correspond to places where the permutations agree. This completes the proof of Lemma 11. □

3.3. *The final construction.* To get the final construction started we use the fact, easily established, that there exists $0 < H < 1$ and $\epsilon > 0$ such that for infinitely many k there is a set $S \subseteq \{0, 1\}^k$ of cardinality 2^{kH} such that each member of S starts with a 1, and any two members of S are at least ϵ apart in $\bar{d}$. We take $\mathcal{S}(1)$ to consist of the subsets of cardinality 1 of such a set S, where $k = k(1)$ is so large that Lemmas 6–11 all hold for $n = |\mathcal{S}(1)| = n(1) = 2^{kH}$ and so that $2^{\sqrt{k}} \leq n \leq 2^{k-1}$. Further requirements on $k(1)$ will be made as we go along. Note that $\mathcal{S}(1)$ is $(\epsilon, 0)$-core-separated.

The next stage $\mathcal{S}(2)$ is defined to be the permutation-separated merging of a strong heap extension $\mathcal{M}(1)$ of $\mathcal{S}(1)$, using $\delta = \delta(1) = (\log n(1))^{-1/2}$. Proceeding by induction, we take $\mathcal{S}(j+1)$ to be the permutation-separated merging of a strong heap extension $\mathcal{M}(j)$ of $\mathcal{S}(j)$, using $\delta = \delta(j) = (\log n(j))^{-1/2}$, where $n(j)$ is the cardinality of $\mathcal{S}(j)$.

Throughout the remainder of this section we use the notation

$$\begin{aligned}
&n(j) = |\mathcal{S}(j)|, \quad \delta(j) = (\log n(j))^{-1/2},\\
&k(j) = \text{length of the sequences that belong to the sets in } \mathcal{S}(j),\\
&b(j) = \lfloor n(j)^{1-\delta(j)} \rfloor = \text{number of sets in an } (\mathcal{S}(j), \delta(j)) \text{ heap},\\
&m(j) = \lfloor \exp(n(j)^{1-3\delta(j)}) \rfloor = \text{number of heaps used to construct } \mathcal{S}(j+1).
\end{aligned}$$

We also have the relations

$$k(j+1) = m(j)(n(j)^2 k(j) + k(j)b(j) + k(j) + 1), \tag{24}$$

$$n(j+1) = \lceil \exp(\delta(j) m(j) \log m(j)) \rceil, \tag{25}$$

and by Lemma 11 we can suppose that $\mathcal{S}(j)$ is $(\epsilon(j), \theta(j))$-core-separated where $\epsilon(1) = \epsilon$, $\theta(1) = 0$, and

$$\epsilon(j+1) = \epsilon(j)(1 - 2\theta(j) - 14\delta(j))(1 - 7\delta(j)), \tag{26}$$

$$\theta(j+1) \leq 2m(j) \exp(-2m(j)\delta(j)^2). \tag{27}$$

The index j will be omitted when it is understood.

It will be important later that $2^{\sqrt{k}} \leq n \leq 2^{k-1}$ holds not only for stage 1, but for all later stages. The lower bound plays a role in showing that more than exponential growth in the number of future measures occurs, while, as already noted, the upper bound makes possible relatively short heap markers. Thus before proceeding further we establish the following lemma.

LEMMA 12. *If $n(1)$ is large enough then $2^{\sqrt{k(j)}} \leq n(j) \leq 2^{k(j)-1}$, $j \geq 1$.*

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

Proof. With $k = k(j), n = n(j), m = m(j)$, and $b = b(j)$ we have

$$n(j+1) \leq m! \leq 2^{m \log m} \leq 2^{mn^2 k} \leq 2^{k(j+1)-1}$$

where we used the fact that $\log m \leq 2n^{1-3\delta} \leq n^2$.

The lower bound has the form

$$k(j) \leq \log^2 n(j), \quad j = 1, 2, 3, \ldots. \tag{28}$$

It was assumed to hold for stage 1, so assume it holds for some $j \geq 1$. We then have

$$\begin{aligned} k(j+1) = m(n^2 k + k + 1 + kb) &\leq m(n^2 \log^2 n + k + 1 + kb) \\ &\leq \frac{m^2 \log^2 m}{4 \log n}, \end{aligned} \tag{29}$$

provided n exceeds some n_0, since $b \leq n^{1-\delta} + 1$ and $m \geq \exp(n^{1-3\delta})$. In particular, the bound (29) holds for all $j \geq 1$ if $n(1) \geq n_0$. The right-hand side of (29) lower bounds $n(j+1)$, by Lemma 10. This shows that (28) holds for all j, provided $n(1)$ is large enough. □

We next establish ergodicity.

LEMMA 13. *The sequence $\{\mathcal{S}(j)\}$ defines an ergodic measure μ on binary sequences.*

Proof. As in our variational construction we show how an iteration stage can be interpreted via cutting and stacking then appeal to general results about such constructions to obtain the desired ergodic measure. (Such results can also be established directly in terms of limiting properties of the ℓ-block distributions in sequences that belong to the members of $\mathcal{S}(j)$ as $j \to \infty$.)

Consider the jth stage $\mathcal{S} = \mathcal{S}(j) = \{A_i : 1 \leq i \leq n(j)\}$, and let $\mathcal{S}(j+1) = \{B_\pi : \pi \in \Pi\}$ be the Π-merging of the extensions of the strong heap separation $\{\mathcal{H}_r^* : r \in [1, m]\}$ of $\mathcal{S}$.

Associated with A_i is a column structure (which we also label by A_i) with $|A_i|$ columns of height k and equal width, the columns being labeled by the sequences that belong to the set A_i. We can regard $\mathcal{S}$ as the union of the column structures $\{A_i\}$. The heaps can then be regarded as column structures by cutting each A_i into copies and assigning one copy to each of the heaps to which A_i belongs. To keep all the heaps of the same size, however, each A_i must be cut into the same number of copies, hence, some mass must be discarded since the number of heaps that contain a copy of A_i depends on i. This mass removal is described rigorously in the following paragraphs.

For each i, define

$$\gamma_i = |\{r : A_i \in \mathcal{H}_r\}|,$$

and put $\gamma = \max_i \gamma_i$. For each i, cut the column structure A_i into γ equal width copies

$$A_{i,1}, A_{i,2}, \ldots, A_{i,\gamma}.$$

Let C_i be the set of all $A_{i,u}$ for which $u \leq \gamma_i$. The column structures of C_i are assigned in a one-to-one manner to those $\mathcal{H}_r$ to which A_i was assigned. In this way, we can think of

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

each $\mathcal{H}_r$ as a column structure consisting of equal-width copies of exactly $b = \lfloor n^{1-\delta} \rfloor$ of the A_i. Note, in particular, that each $\mathcal{H}_r$ has the same mass.

For each i, the set of all $A_{i,u}$ for which $\gamma_i < u \leq \gamma$ is discarded. The fraction of mass discarded from A_i is $1 - \gamma_i/\gamma$, which, by Lemma 6(b), satisfies the bound

$$1 - \frac{\gamma_i}{\gamma} \leq 1 - \frac{mn^{-\delta} - m^{3/4}n^{-\delta/2}}{mn^{-\delta} + m^{3/4}n^{-\delta/2}} \leq 2m^{-1/4}n^{\delta/2}. \tag{30}$$

The t-extension $\mathcal{H}_r^t$ can be regarded as the t-fold independent cutting and stacking of $\mathcal{H}_r$. The heap marker $h(\mathcal{H}_r)$ can be regarded as a single column of height $k + 1 + kb$ and width that of $\mathcal{H}_r^t$, labeled by the marker of $\mathcal{H}_r$; thus $\mathcal{H}_r^*$ is now just the stacking of $\mathcal{H}_r^t$ on top of $h(\mathcal{H}_r)$. The mergings $\{B_\pi : \pi \in \Pi\}$ are obtained by cutting each $\mathcal{H}_r^*$ into $N = |\Pi|$ equal width copies, $\{F_{r,\pi} : \pi \in \Pi\}$, and taking B_π to be the successive independent cutting and stacking of $\{F_{\pi(i),\pi} : 1 \leq i \leq m\}$ in order of increasing i.

The mass-removal bound (30) is summable in the stage index j, since once j is large enough to guarantee that $1 - 3\delta(j) \geq 1/2$, it is bounded above by $2\exp(-\frac{1}{4}\sqrt{n})2^{(1/2)\sqrt{\log n}}$, which is summable in n. Furthermore, removal at latter stages does not affect the relative distribution of the remaining mass at earlier stages. Thus, after removal of this mass at each stage, the entire picture can be normalized so the total amount of mass is exactly 1 in the limit. In other words, after normalizing we can suppose

$$\lambda(\mathcal{S}(j)) \to 1, \quad \text{as } j \to \infty,$$

where λ denotes the Lebesgue measure on the unit interval. Thus, since the total mass of the tops of the jth stage structure shrinks to 0 as $j \to \infty$, the sequence $\{\mathcal{S}(j)\}$ defines a stationary measure μ on binary sequences.

As in the variational case, to establish ergodicity it is enough to show that Lemma 1 holds. To establish this we again use the notation $\tilde{x}_1^{tk}$ to denote the sequence obtained from $x_1^K \in B_\pi$ by deleting the markers of the $\mathcal{H}_{\pi(i)}^*$ and write

$$\tilde{x}_1^{mtk} = u_1 u_2 \cdots u_m, \quad u_i \in \mathcal{H}_{\pi(i)}^t, \quad 1 \leq i \leq m.$$

Each u_i can then be expressed as a k-block concatenation

$$u_i = b(1,i)b(2,i)\cdots b(t,i), \quad b(j,i) \in \mathcal{H}_{\pi(i)}, \quad 1 \leq j \leq t.$$

Fix $\ell \leq n$ and $a_1^k \in A_\ell$, and let $Y = Y(a_1^k | x_1^K)$ denote the number of times there is a pair (j,i) such that $b(j,i) = a_1^k$. Exactly γ_ℓ of the $\mathcal{H}_r$ contain a copy of A_ℓ, each $\mathcal{H}_r$ has cardinality $\lfloor n^{1-\delta} \rfloor$, and independent cutting and stacking is used to form the mergings, hence, conditioned on the union of the $\mathcal{H}_r$ that contain A_ℓ, Y is a binomial

$$\left(t\gamma_\ell, \frac{1}{\lfloor n^{1-\delta} \rfloor |A_\ell|} \right)$$

random variable. Note, however, that $t\gamma_\ell \gg \lfloor n^{1-\delta} \rfloor |A_\ell|$ since $t = n^2$ and γ_ℓ is approximately $mn^{-\delta}$, by Lemma 6(b). Thus the argument used to prove Lemma 1 extends to this case, so that μ is indeed ergodic. This completes the proof of Lemma 13. □

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

3.4. *Super-exponential growth.* It remains to show that by making $k(1)$ larger, if necessary, we obtain a measure μ for which the super-exponential growth property (2) holds. Towards this end we first establish a $\bar{d}$ analogue of the variational separation property expressed in Lemma 2. In our new setting the height of a heap extension $\mathcal{H}_i^*$ is $h = n^2k + kb + k + 1$. As before, for L in the range

$$kb + k + 1 < L \leq h - 2k,$$

we let $[L]$ denote the union of the intervals that lie at level L of the columns of the heap extensions $\mathcal{H}_i^*$, for $x \in [L]$, we define $r(x) = r$ if x lies in the support of $\mathcal{H}_r^*$, and we note that r is completely determined by the past x_{-L+1}^0.

LEMMA 14. *If $x, y \in [L]$ and $r(x) \neq r(y)$, then*

$$\bar{d}_{2k}(\mu_{2k}(\cdot|x_{-\infty}^0), \quad \mu_{2k}(\cdot|y_{-\infty}^0)) \geq \epsilon(1 - 2\theta - 4\delta)/2.$$

Proof. As before, we let s be the least non-negative integer such that $L + s - (kb + k + 1)$ is a multiple of k, and define the marginal μ_{2k}^s of a measure μ_{2k} on A^{2k} by the formula

$$\mu_{2k}^s(a_{s+1}^{s+k}) = \sum_{a_1^s, a_{s+k+1}^{2k}} \mu_{2k}(a_1^{2k}).$$

We then have, see (12), the uniform distribution property

$$\mu_{2k}^s(a_{s+1}^{s+k}|x_{-\infty}^0) = \begin{cases} |\mathcal{H}_{r(x)}|^{-1}, & \text{if } a_{s+1}^{s+k} \in \mathcal{H}_{r(x)}, \\ 0, & \text{otherwise.} \end{cases}$$

The lemma now follows from Lemma 7 and the $\bar{d}$ contraction property, i.e. $\bar{d}_{2k}(\mu_{2k}, \nu_{2k}) \geq (1/2)\bar{d}_k(\mu_{2k}^s, \nu_{2k}^s)$. □

Lemma 5 now takes the following form.

LEMMA 15. *Given a past measurable subset $B \subset \{0, 1\}^Z$ of measure $\gamma < 1/2$, for all sufficiently large j there is a finite set G, disjoint from B, such that:*

(i) $|G| \geq (1 - 2\gamma)\exp(n^{1-3\delta})$; *and*

(ii) $\bar{d}_{2k}(\mu_{2k}(\cdot|x_{-\infty}^0), \mu_{2k}(\cdot|y_{-\infty}^0)) \geq \epsilon(j)(1 - 2\theta - 4\delta)/2, x, y \in G, x \neq y$, *where $n = n(j), \delta = \delta(j), k = k(j)$, and $\theta = \theta(j)$.*

Proof. The proof is similar to the proof of Lemma 5, but easier, since heaps overlap by at most $2n^{-\delta}|\mathcal{H}|$, see Lemma 6(a). □

It remains to show that if $k(1)$ is large enough then both of the following hold:

(a) $\inf_j \epsilon(j)(1 - 2\theta(j) - 4\delta(j)) > 0$;

(b) $\exp(n(j)^{1-3\delta(j)})$ is super-exponential in $2k(j)$.

We assume $k(1)$ is large enough to guarantee that $1 - 3\delta(1) > 1/2$ and hence that $1 - 3\delta(j) > 1/2$ for all $j \geq 1$. To establish (a) it is enough to show that both $\sum \theta(j)$ and $\sum \delta(j)$ are finite. The bound (27) gives

$$\theta(j + 1) \leq 2m \exp(-2m\delta^2)$$

$$\leq 2^{n+1} \exp\left(-2\frac{2^{\sqrt{n}}}{\log n}\right),$$

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

which is summable in n, hence in j. Likewise, the bound (25) gives

$$\frac{1}{\sqrt{\log n(j+1)}} \le \frac{2}{\delta m \log m} \le \frac{2\sqrt{\log n}}{2^{\sqrt{n}}\sqrt{n}}$$

which is also summable in n.

Next we use $k \le (\log n)^2$, see Lemma 12, and $\log m \ge \sqrt{n}$ to obtain

$$\frac{\log m(j)}{k(j)} \ge \frac{\sqrt{n}}{k} \ge \frac{\sqrt{n}}{\log^2 n}$$

which goes to infinity as $n \to \infty$.

We have shown that our construction has the desired properties, hence the proof of Theorem 2 is now complete. □

Acknowledgements. We are grateful to Benjy Weiss and Steve Kalikow for many useful conversations on the topic of this paper. We especially wish to thank the referee who pointed out many errors and obscurities in the first two submissions.

Katalin Marton was partially supported by Hungarian National Foundation for Scientific Research Grants T026041 and T016386 and by MTA-NSF project 37, and Paul C. Shields was partially supported by joint NSF–Hungarian Academy grant INT-9515485.

A. *Appendix*

To simplify the notation in the following proofs we take m and $b = n^{1-\delta}$ to be integers, ignoring that fact that rounding to the nearest integer may be required.

Proof of Lemma 3. Robbins' improvement of Stirling's formula, see [**1**, p. 54],

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12(n+1))} \le n! \le \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n)}, \tag{A.1}$$

immediately yields the lower bound in

$$\frac{\exp(nh(k/n))}{Kn} \le \binom{n}{k} \le \exp(nh(k/n)), \tag{A.2}$$

where $h(t) = -t\ln t - (1-t)\ln(1-t)$ is the so-called binary entropy function (expressed in terms of natural logarithms), exp denotes the exponent with base e (rather than base 2 as earlier), and K is a positive constant, independent of n and k. The upper bound is well known. Also useful are the bounds

$$-\alpha\ln\alpha \le h(\alpha) \le -\alpha\ln\alpha + \alpha, \quad 0 < \alpha < 1. \tag{A.3}$$

The number of heaps is the number of combinations of n objects taken $n^{1-\delta}$ at a time, for which the left-hand inequality in (A.2) yields the lower bound

$$\binom{n}{n^{1-\delta}} \ge \exp(nh(n^{-\delta}) - \ln Kn). \tag{A.4}$$

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

For a given heap $\mathcal{H}$, the number of heaps $\mathcal{H}_1$ for which $|\mathcal{H} \cap \mathcal{H}_1| > \delta n^{1-\delta}$ is bounded above by the product

$$n \binom{n^{1-\delta}}{\delta n^{1-\delta}} \binom{n}{(1-\delta)n^{1-\delta}},$$

which, using the right-hand inequality in (A.2), is bounded above by

$$n \exp(n^{1-\delta} h(\delta) + nh((1-\delta)n^{-\delta})). \tag{A.5}$$

A greedy algorithm that starts with a heap $\mathcal{H}_1$ and inductively selects a heap $\mathcal{H}_{i+1}$ such that $|\mathcal{H}_{i+1} \cap \mathcal{H}_j| \leq \delta n^{1-\delta}$, for all $j \leq i$, produces at least the ratio of lower bound (A.4) to upper bound (A.5) pairwise $(1-\delta)$-separated heaps, and hence

$$C_n^*(\delta) \geq \exp(nh(n^{-\delta}) - n^{1-\delta} h(\delta) - nh((1-\delta)n^{-\delta}) + O(\ln n)).$$

An application to this of the entropy bounds (A.3) followed by some algebra yields the bound

$$C_n^*(\delta) \geq \exp(n^{1-\delta}\delta^2 \log n - O(n^{1-\delta}) - O(\ln n)).$$

For $\delta = C(\log n)^{-1/2})$ and C large enough the first term in the exponent dominates the remaining terms. This proves Lemma 3. □

Proof of Lemma 6. We leave it to the reader to show that a direct application of the preceding proof yields a bound on the exponent that goes to zero, thus a subtler approach is needed. Towards this end we establish the following lemma.

LEMMA A.1. *If $b = \epsilon n$, $\epsilon = n^{-\delta}$, and $\gamma = 2n^{-\delta}$, then*

$$\frac{1}{n} \ln \frac{\binom{b}{\gamma b}\binom{n-b}{b-\gamma b}}{\binom{n}{b}} \leq \frac{\epsilon(\gamma-\epsilon)^2}{2\gamma(1-\gamma)(1-\epsilon)} + O\left(\frac{\ln b}{b}\right). \tag{A.6}$$

Proof. Put

$$p = \epsilon, \quad p_1 = \gamma, \quad p_2 = \epsilon \frac{1-\gamma}{1-\epsilon}$$

and note that $p = \epsilon p_1 + (1-\epsilon)p_2$. Concavity of entropy, the fact that $h''(p) = -1/p(1-p)$, and the Taylor approximation give

$$\begin{aligned} h(p) - \epsilon h(p_1) - (1-\epsilon)h(p_2) &\geq \frac{1}{2} \min_{p_2 \leq \eta \leq p_1} |h''(\eta)| \epsilon(1-\epsilon)(p_1 - p_2)^2 \\ &= \frac{1}{2} \frac{\epsilon(\gamma-\epsilon)^2}{\gamma(1-\gamma)(1-\epsilon)}. \end{aligned}$$

This, together with the binomial bounds (A.2), then yields the lemma. □

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

COROLLARY A.1. *If we select two subsets of size b from a set with n elements at random according to the uniform distribution on b element subsets, the probability of the event that they intersect in at least γb elements is bounded above by*

$$b \exp\left(-n\left[\frac{\epsilon(\gamma-\epsilon)^2}{2\gamma(1-\gamma)(1-\epsilon)} - O\left(\frac{\ln b}{b}\right)\right]\right), \tag{A.7}$$

which is at most $-\frac{1}{8}n^{1-2\delta}$, for n large enough.

Proof. If γb is an integer (which we assume) then the probability that the intersection has exactly γb elements is

$$\frac{\binom{b}{\gamma b}\binom{n-b}{b-\gamma b}}{\binom{n}{b}}. \tag{A.8}$$

The expected size of the intersection is ϵb, since each element in the first subset is selected into the second one with probability $b/n = \epsilon$. Since $\gamma \geq \epsilon$, the probability that the intersection contains $\ell \geq \gamma b$ elements is overbounded by b times the quantity in (A.8). An application of Lemma A.1 gives the desired bound (A.7). □

Now we are ready to prove Lemma 3. We randomly select $m = \lceil \exp(n^{1-3\delta}) \rceil$ subsets of an n-set of size b. The corollary implies that with probability approaching 1 as $n \to \infty$ there will be a collection $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_m\}$ of $(\mathcal{S}, \delta)$ heaps such that

(a) $|\mathcal{H}_r \cap \mathcal{H}_s| \leq 2n^{1-2\delta}, r \neq s.$

The probability that a given index i belongs to a given $\mathcal{H}_r$ is $n^{-\delta}$, so the number of heaps to which i belongs is a binomial (m, p) random variable, with $p = n^{-\delta}$. Thus the expected number of heaps to which i belongs is $mn^{-\delta}$, and the probability that

$$||\{r : i \in \mathcal{H}_r\}| - mn^{-\delta}| \geq m^{3/4}\sqrt{n^{-\delta}}$$

goes to 0 as $n \to \infty$. Thus

(b) $||\{r : A_i \in \mathcal{H}_r\}| - mn^{-\delta}| \leq m^{3/4}\sqrt{n^{-\delta}}, i \in \mathcal{S}$

holds with probability approaching 1, as $n \to \infty$.

In summary, if n is large enough, there must be a collection of m subsets of $\mathcal{S}$ of size $n^{1-\delta}$ for which (a) and (b) both hold. □

Proof of Lemma 10. Given $\gamma > 0$ and m, let $N = N(m, \gamma)$ be the maximum cardinality of a collection of pairwise $(1-\gamma)$-separated m-permutations. To prove the lemma it is enough to show that

$$\log N \geq \gamma m \log m + m\gamma \log \frac{\gamma}{e} + \frac{1}{2}\log \frac{\gamma}{m}. \tag{A.9}$$

There are, however, at most

$$\binom{m}{j}(m-j)! = \frac{m!}{j!} \tag{A.10}$$

m-permutations that agree with a given one at j places. If $j \geq \delta m$ then (A.10) is bounded above by $m!/(m\delta)!$. Since there at most m such j's, a greedy algorithm yields $N \geq (m\delta)!/m$, to which an application of Robbins' bound, (A.1), establishes the lemma. □

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press

REFERENCES

[1] W. Feller. *An Introduction to Probability Theory and its Applications*, Vol. I, 2nd edn. Wiley, New York, 1971.

[2] S. Kalikow. T, T^{-1} transformation is not loosely Bernoulli. *Ann. Math.* **115** (1982), 393–409.

[3] D. Ornstein and P. Shields. The $\bar{d}$-recognition of processes. *Adv. Math.* **104** (1994), 182–224.

[4] D. Ornstein and P. Shields. An uncountable family of K-automorphisms. *Adv. Math.* **10** (1973), 63–88.

[5] D. Ornstein and B. Weiss. How sampling reveals a process. *Ann. Prob.* **18** (1990), 905–930.

[6] P. C. Shields. *The Ergodic Theory of Discrete Sample Paths (AMS Graduate Studies in Mathematics)*. American Mathematics Society, Providence, RI, 1996.

[7] J.-P. Thouvenot. On the stability of the weak Pinsker property. *Israel J. Math.* **7** (1977), 150–162.

https://doi.org/10.1017/S0143385702000123 Published online by Cambridge University Press