ISRAEL JOURNAL OF MATHEMATICS 91 (1995), 215–237

RANDOMLY GENERATED DISTRIBUTIONS

BY

R. DANIEL MAULDIN* AND MICHAEL G. MONTICINO

Mathematics Department, University of North Texas
P.O. Box 5116, Denton, Texas 76203-51116, USA
e-mail: mauldin@unt.edu mgmont@unt.edu

ABSTRACT

A new scheme for randomly generating probability distributions on the interval $[0,1]$ is introduced. The scheme can also be viewed as a way to generate homeomorphisms at random. Conditions are given so that a continuous measure with full support is generated almost surely. Geometric properties of the generated probability measures are examined, including the dimension and derivative structure of the measures and their respective distribution functions. For example, we give conditions so that almost all the distribution functions of the measures generated are strictly singular. Applications include determining average case errors for numerical methods of equation solving and Bayesian statistics.

1. Introduction

This paper presents a new scheme for randomly generating probability measures on [0,1] — that is, for obtaining probability measures or priors on the space of probability measures on [0,1]. The scheme can also be viewed as a way to generate a homeomorphism at random. From this latter perspective, we generalize Graf, Mauldin and Williams' (1986) method of randomly generating homeomorphisms. Other related methods of generating probability measures and homeomorphisms are described by Dubins and Freedman (1967) and by Kraft (1964).

One application of the priors we obtain is in determining average case errors for numerical methods of equation solving. As mentioned in Novak (1988), worst case errors for numerical methods are generally much larger than those encountered

* Research supported by NSF Grant DMS-9303888.
Received March 4, 1993 and in revised form December 28, 1993

with most functions. Therefore, there is interest in finding average error bounds. Such average case errors are investigated by Graf, Novak, and Papageorgiou (1989) and Ritter (1992). Novak (1989) and Novak and Ritter (1992) survey related results on average case errors. Other applications are discussed by Ulam (1982), Graf, Mauldin and Williams (1986) and Mauldin, Sudderth and Williams (1992). In particular, the priors produced here — random rescaling priors — are tailfree priors as described in Ferguson (1973) and Doksum (1974). A related set of priors are the Dirichlet priors given by Ferguson (1974). Random rescaling priors, unlike Dirichlet priors, can give probability one to the set of continuous measures with full support. Such priors are useful in Bayesian statistics.

In the second section, we introduce our scheme and mention related work. Conditions so that a continuous probability measure with full support is generated almost surely are given in section 3. Equivalently, as discussed in the next section, these conditions ensure that the scheme generates a homeomorphism almost surely.

Sections 4 and 5 examine the geometric properties of the generated probability measures. Among the properties we consider are the derivative structure of the distributions and the (Hausdorff) dimension of the supports.

Specifically, Theorem 4.1 gives a condition under which almost all distribution functions are strictly singular — i.e., do not have a finite positive derivative anywhere. And, in Theorem 4.2, we establish conditions under which, for each $x \in [0,1]$, almost all distributions have derivative 0 at x. One of these conditions involves bounding the fourth moments of a particular set of random variables. We believe this condition can be weakened to only bounding the second moments, but we have been unable to prove it. Of course, Theorem 4.2 implies Theorem 4.1. An example is given which shows that the converse does not hold.

Upper and lower bounds on the dimensions of the generated probability measures are derived in section 5. The martingale convergence theorem and a Frostman type lemma are used to obtain the bounds.

2. Random rescaling

In this section, we describe our scheme for randomly generating a probability measure. First, note that a distribution function of a probability measure on [0,1] is a non-decreasing right continuous function from [0,1] to [0,1], which assumes the value 0 at 0 and the value 1 at 1. If the probability measure is continuous and has full support, then its distribution function will be continuous and strictly increasing — that is, a homeomorphism of [0,1] onto [0,1]. On the other hand, a non-decreasing right continuous function from [0,1] to [0,1] which leaves 0 and 1 fixed determines a unique probability measure on [0,1]. If the function is also a homeomorphism of [0,1] onto [0,1], then the induced probability measure will be continuous and have full support. Thus, generating a probability measure at random is equivalent to selecting a non-decreasing right continuous function at random — a theme initiated by Dubins and Freedman (1967) — and generating a continuous probability measure with full support is equivalent to selecting a homeomorphism at random.

Denote the set of probability measures on the interval [0,1] by $\mathcal{P}([0,1])$. Let τ be a mapping (transition kernel) from the dyadic rationals, $\mathcal{D}$, to $\mathcal{P}([0,1])$. We generate a distribution function, h, of a probability measure on [0,1] by **randomly rescaling** τ as follows. First, set the value of h at 0, $h(0)$, to 0 and set $h(1) = 1$. Now randomly select the value of $h(\frac{1}{2})$ according to the distribution of $\tau(\frac{1}{2})$. Select $h(\frac{1}{4})$ according to the distribution of $\tau(\frac{1}{4})$ scaled to the interval $[0, h(\frac{1}{2})]$ and independently select $h(\frac{3}{4})$ according to $\tau(\frac{3}{4})$ scaled to $[h(\frac{1}{2}), 1]$. Continue in this manner to define a function on the dyadic rationals. In the natural way, extend h to a distribution function of a probability measure on [0,1]. This scheme thus induces a probability measure or prior, denoted by R_τ, on the space of distribution functions — or, equivalently, on $\mathcal{P}([0,1])$.

A more precise view of R_τ can be obtained by introducing a **scaling map** θ from $[0,1]^{\mathcal{D}}$ to $\mathcal{P}([0,1])$. (Note that we alternately regard images of θ as distribution functions and as the corresponding probability measures.) Let $\mathcal{D}_n$ be the set of strictly nth level dyadic rationals — e.g., $\frac{1}{2} \notin \mathcal{D}_2$. For $t = (t(\frac{1}{2}), t(\frac{1}{4}), t(\frac{3}{4}), \ldots) \in [0,1]^{\mathcal{D}}$, define the distribution function $\theta(t)$ inductively on $\mathcal{D}$ as follows:

$$\theta(t)(0) = 0, \quad \theta(t)(1) = 1.$$

Assume $\theta(t)|_{\bigcup_{i=1}^{n} \mathcal{D}_i}$ has been defined. And, for $0 \leq j \leq 2^n - 1$, set

$$\theta(t)\left(\frac{2j+1}{2^{n+1}}\right) = \theta(t)\left(\frac{j}{2^n}\right) + \left(\theta(t)\left(\frac{j+1}{2^n}\right) - \theta(t)\left(\frac{j}{2^n}\right)\right) t\left(\frac{2j+1}{2^{n+1}}\right).$$

It is straightforward to show θ is a well defined, open and continuous map from $[0,1]^{\mathcal{D}}$ into $\mathcal{P}([0,1])$. The prior R_τ is induced by $P_\tau = \prod_{d \in \mathcal{D}} \tau(d)$ through θ.

As mentioned above, Graf, Mauldin and Williams (1986) and Dubins and Freedman (1967) introduced related methods of generating probability measures and homeomorphisms. In particular, one scheme presented by Graf, Mauldin, and Williams (1986) randomly rescales a fixed measure μ at each stage — that is, their transition kernel equals μ at each dyadic rational. So, obviously, our scheme generalizes their method and we produce a larger class of priors than they do. To see that random rescaling can also produce priors which can not be constructed within the framework of Dubins and Freedman (1967), set $\tau(d) = \delta_{\{\frac{1}{3}\}}$, for all dyadic rationals $d < \frac{1}{2}$, and set $\tau(d) = \delta_{\{\frac{1}{2}\}}$, for all $d \geq \frac{1}{2}$. Then R_τ, regarded as a probability measure on distribution functions, is $\delta_{\{\tilde{h}\}}$, where $\tilde{h}(x) = x$, for $x \geq \frac{1}{2}$, and $\tilde{h}$ is singular over $(0, \frac{1}{2})$ (see, for instance, section 4). Dubins and Freedman (1967, Theorem 5.1) show that this prior can not be obtained from their construction for any basic measure $\mu \in \mathcal{P}([0,1]^2)$. More generally, for any homeomorphism, h, there exists a transition kernel τ for which $R_\tau = \delta_{\{h\}}$. However, if h is not singular almost everywhere or if h is not the identity function, then $\delta_{\{h\}}$ can not be obtained from a Dubins and Freedman (1967) construction.

3. Continuous measures with full support

Here general conditions are determined so that the random rescaling scheme generates continuous measures with full support almost surely — or, equivalently, conditions so that a homeomorphism is generated almost surely.

Recall that a probability measure $\pi \in \mathcal{P}([0,1])$ is **continuous** if $\pi(\{x\}) = 0$, for all $x \in [0,1]$. Let

$$\mathcal{C} = \{\pi \in \mathcal{P}([0,1]) \colon \pi(\{x\}) = 0, \text{ for all } x \in [0,1]\}.$$

A probability measure μ defined on a compact Hausdorff space $\mathcal{M}$ has **full support** if every nonempty open subset of $\mathcal{M}$ has positive μ-measure. Note that this is equivalent to $\mathcal{M}$ being the smallest compact set which has μ-measure one. Let

$$\mathcal{S} = \{\pi \in \mathcal{P}([0,1]) \colon \pi \text{ has full support}\}$$

and

$$\tilde{\mathcal{C}} = \mathcal{C} \cap \mathcal{S}.$$

To establish conditions on a transition kernel $\tau \colon \mathcal{D} \to \mathcal{P}([0,1])$ so that $R_\tau(\mathcal{C}) = 1$, we borrow a technique from Mauldin, Sudderth, and Williams (1992). Suppose R_τ is a random rescaling prior and let Π_τ be a random measure with distribution R_τ. Then there exists a unique (distribution-wise) sequence of [0,1]-valued random variables $X_1^\tau, X_2^\tau, \ldots$ which, given Π_τ, are independent each with distribution Π_τ. That is, for $X^\tau = (X_1^\tau, X_2^\tau, \ldots)$ and Borel set $A \subseteq [0,1] \times [0,1] \times \ldots$,

$$P[X^\tau \in A] = \int \pi^\infty(A)\, dR_\tau(\pi), \tag{3.1}$$

where $\pi^\infty = \pi \times \pi \times \cdots$ is the infinite product measure on $[0,1] \times [0,1] \times \cdots$. (This sequence is an exchangeable sequence of random variables directed by R_τ. See, for instance, Aldous (1983) for details.) The sequence $X_1^\tau, X_2^\tau, \ldots$ gives us a way to check whether $R_\tau(\mathcal{C}) = 1$. In particular, we get Proposition 3.1. The proposition is essentially Lemma 5.2 of Mauldin, Sudderth, and Williams (1992).

PROPOSITION 3.1: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$. Then $R_\tau(\mathcal{C}) = 1$ if and only if $P[X_1^\tau = X_2^\tau] = 0$.*

Proof: From (3.1),

$$\begin{aligned} P[X_1^\tau = X_2^\tau] &= \int \int_{[0,1]} \pi(\{x\}) d\pi(x) dR_\tau(\pi) \\ &= \int_{\mathcal{C}} \int_{[0,1]} \pi(\{x\}) d\pi(x) dR_\tau(\pi) + \int_{\mathcal{C}^c} \int_{[0,1]} \pi(\{x\}) d\pi(x) dR_\tau(\pi) \\ &= 0 + \int_{\mathcal{C}^c} \int_{[0,1]} \pi(\{x\}) d\pi(x) dR_\tau(\pi) \\ &> 0 \end{aligned}$$

if and only if $R_\tau(\mathcal{C}^c) > 0$. ∎

We say a transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$ is **centered** if, for each $\epsilon > 0$, there exists a $\delta \in (0, \frac{1}{2})$, such that

$$\tau(d)\Big((\delta, 1-\delta)\Big) > 1 - \epsilon,$$

for all $d \in \mathcal{D}$.

Denote the distribution function of a probability measure π on [0,1] by h_π.

THEOREM 3.2: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$. If τ is centered, then $R_\tau(\mathcal{C}) = 1$.*

Proof: It suffices to show that $P[X_1^\tau = X_2^\tau] = 0$. Let

$$E = \{(x_1, x_2) \in [0,1]^2 \colon x_1 = x_2\}.$$

For each $n = 1, 2, \dots$ and $2 \le i \le 2^n$, let $E_{1,n} = \left[0, \frac{1}{2^n}\right]$ and $E_{i,n} = \left(\frac{i-1}{2^n}, \frac{i}{2^n}\right]$.

From (3.1), for every $n = 1, 2, \dots,$

$$\begin{aligned} P[X_1^\tau = X_2^\tau] &= \int \pi^2(E) dR_\tau(\pi) \\ &\le \int \pi^2\Big(\bigcup_{i=1}^{2^n} E_{i,n}^2\Big) dR_\tau(\pi). \end{aligned}$$

By construction,

$$\begin{aligned} \int \pi^2(E_{1,1}^2 \cup E_{2,1}^2) dR_\tau(\pi) &= \int \left(h_\pi\left(\frac{1}{2}\right)\right)^2 + \left(1 - h_\pi\left(\frac{1}{2}\right)\right)^2 dR_\tau(\pi) \\ &= \int y^2 + (1-y)^2 d\tau(\frac{1}{2})(y). \end{aligned}$$

More generally, if $p_{i,n} = \int \pi^2(E_{i,n}^2) dR_\tau(\pi)$, then

$$\int \pi^2(E_{2i-1,n+1}^2 \cup E_{2i,n+1}^2) dR_\tau(\pi) = p_{i,n} \int y^2 + (1-y)^2 d\tau\left(\frac{2i-1}{2^{n+1}}\right)(y)$$

Now let $\bar{\delta} \in (0, \frac{1}{2})$ be such that, for all $d \in \mathcal{D}$,

$$\tau(d)((\bar{\delta}, 1 - \bar{\delta})) > \frac{1}{2}.$$

So, for $d \in \mathcal{D}$,

$$\begin{aligned} \int y^2 + (1-y)^2 d\tau(d)(y) &= 1 + 2\left(\int y^2 d\tau(d)(y) - \int y d\tau(d)(y)\right) \\ &= 1 + 2\left(\int_{[0,1-\bar{\delta}]} y^2 d\tau(d)(y) + \int_{(1-\bar{\delta},1]} y^2 d\tau(d)(y)\right. \\ &\qquad \left. - \int y d\tau(d)(y)\right) \\ &\le 1 + 2\left((1-\bar{\delta}) \int_{[0,1-\bar{\delta}]} y d\tau(d)(y) + \int_{(1-\bar{\delta},1]} y d\tau(d)(y)\right. \\ &\qquad \left. - \int y d\tau(d)(y)\right) \\ &= 1 - 2\bar{\delta}\left(\int_{[0,1-\bar{\delta}]} y d\tau(d)(y)\right) \\ &= 1 - 2\bar{\delta}\left(\int_{[0,\bar{\delta}]} y d\tau(d)(y) + \int_{(\bar{\delta},1-\bar{\delta}]} y d\tau(d)(y)\right) \\ &\le 1 - \bar{\delta}^2 \end{aligned}$$

Hence, letting $\rho = 1 - \bar{\delta}^2 < 1$, an easy induction argument shows

$$\int \pi^2\Big(\bigcup_{i=1}^{2^n} E_{i,n}^2\Big) dR_\tau(\pi) \le \rho^n$$

for each $n = 1, 2, \ldots$. And so $P[X_1^\tau = X_2^\tau] = 0$. ∎

To ensure that $R_\tau(\mathcal{S}) = 1$ we need only that $\tau(d)(\{0,1\}) = 0$, for each $d \in \mathcal{D}$. We state this formally in Theorem 3.3.

THEOREM 3.3: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$. $R_\tau(\mathcal{S}) = 1$ if and only if $\tau(d)(\{0,1\}) = 0$, for each $d \in \mathcal{D}$.*

Proof: Consider the reverse implication first. Note that

$$\mathcal{S}^c = \bigcup_{n=1}^{\infty} \bigcup_{i=1}^{2^n} S_{i,n},$$

where $S_{i,n} = \{\pi\colon \pi(E_{i,n}) = 0\}$ and $E_{i,n}$ is as defined in the proof of Theorem 3.2. By hypothesis,

$$\begin{aligned} R_\tau(S_{1,1}) &= R_\tau(\{\pi\colon \pi([0, \tfrac{1}{2}]) = 0\}) \\ &= R_\tau(\{\pi\colon h_\pi(\tfrac{1}{2}) = 0\}) = \tau(\tfrac{1}{2})(\{0\}) = 0. \end{aligned}$$

Similarly,

$$\begin{aligned} R_\tau(S_{2,1}) &= R_\tau(\{\pi\colon \pi((\tfrac{1}{2}, 1]) = 0\}) \\ &= R_\tau(\{\pi\colon h_\pi(\tfrac{1}{2}) = 1\}) = \tau(\tfrac{1}{2})(\{1\}) = 0. \end{aligned}$$

Suppose that $R_\tau(S_{i,n}) = 0$ for all $1 \le i \le 2^n$ and $n \le m$. And, let $1 \le i \le 2^m$. Then

$$\begin{aligned} R_\tau(S_{2i-1,m+1}) = R_\tau(&\{\pi\colon \pi(E_{i,m}) = 0\}) \\ &+ R_\tau(\{\pi\colon \pi(E_{i,m}) \ne 0 \text{ and } \pi(E_{2i-1,m+1}) = 0\}) \\ = 0 &+ R_\tau(\{\pi\colon \pi(E_{i,m}) \ne 0\}) \cdot \tau\left(\frac{2i-1}{2^{m+1}}\right)(\{0\}) \\ = 0&. \end{aligned}$$

The second to last equality follows from the induction assumption and by our scheme for generating π. The last equality holds by hypothesis. Similarly,

$$\begin{aligned} R_\tau(S_{2i,m+1}) &= R_\tau\Big(\{\pi\colon \pi(E_{i,m})=0\}\Big) + R_\tau\Big(\{\pi\colon \pi(E_{i,m})\neq 0 \\ &\qquad\qquad\qquad\qquad \text{and } \pi(E_{2i,m+1})=0\}\Big) \\ &= 0 + R_\tau\Big(\{\pi\colon \pi(E_{i,m})\neq 0\}\Big)\cdot\tau\left(\frac{2i-1}{2^{m+1}}\right)(\{1\}) \\ &= 0. \end{aligned}$$

Hence, by induction, $R_\tau(S_{i,n}) = 0$, for all n and $1 \le i \le 2^n$. Therefore, $R_\tau(\mathcal{S}) = 1$.

The argument establishing the forward implication is analogous to that given for the reverse. ∎

The following corollary is an immediate consequence of Theorems 3.2 and 3.3.

COROLLARY 3.4: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$. If τ is centered, then $R_\tau(\tilde{C}) = 1$.*

As might be expected, several results relating the support of R_τ and the supports' of the $\tau(d)'s$ can be given. We state three such results in Theorem 3.5. For instance, the first result states when R_τ has full support. The proof of the theorem is straightforward and in the interest of space we forego presenting it. We just remark that the theorem can be proved by utilizing the relationship between R_τ and P_τ and standard facts about product spaces. Also, assertion (i) of the theorem can be proved using methods similar to those applied in the proof of Theorem 6.1 of Mauldin, Sudderth, and Williams (1992).

THEOREM 3.5: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$.*

(i) *R_τ has full support on $\mathcal{P}([0,1])$ if and only if, for all $d \in \mathcal{D}$, $\tau(d)$ has full support on $[0,1]$.*
(ii) *The support of R_τ is nowhere dense if and only if either the support of $\tau(d)$ is nowhere dense, for some $d \in \mathcal{D}$, or there exist infinitely many $d \in \mathcal{D}$ for which $\tau(d)$ does not have full support.*
(iii) *If $R_\tau(\tilde{C}) = 1$, then the support of R_τ is dense in itself if and only if either the support of $\tau(d)$ is dense in itself, for some $d \in \mathcal{D}$, or there exist infinitely many $d \in \mathcal{D}$ with $\tau(d)$ not supported by a single point.*

4. Derivatives of the generated distribution functions

Here we examine the derivative structure of the distribution functions generated by random rescaling. A distribution function is **strictly singular** if it does not have a finite positive derivative anywhere. The two main results of this section are the following theorems:

THEOREM 4.1: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$. Suppose that there exists a $\delta > 0$ and a compact set $K \subseteq [0,1] - \{\frac{1}{2}\}$ such that*

$$\tau(d)(K) > \delta,$$

for all $d \in \mathcal{D}$. Then R_τ-almost all h_π are strictly singular.

THEOREM 4.2: *Let R_τ be the random rescaling prior associated with transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$. Suppose that there exists real constants $a_1, a_2 > 0$ such that, for every $d \in \mathcal{D}$,*

$$\begin{aligned} &\int_{(0,1)} \ln(y) d\tau(d)(y) < -\ln 2 - a_1, \\ &\int_{(0,1)} \ln(1-y) d\tau(d)(y) < -\ln 2 - a_1, \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} &\int_{(0,1)} (\ln(y))^4 d\tau(d)(y) < -\ln 2 - a_2, \\ &\int_{(0,1)} (\ln(1-y))^4 d\tau(d)(y) < -\ln 2 - a_2. \end{aligned} \tag{4.2}$$

Then, for every $x \in [0,1]$, R_τ-almost all h_π have derivative 0 at x.

Theorem 4.2, of course, implies that R_τ-almost all h_π are singular. In fact, conditions (4.1) and (4.2) imply the conditions given in Theorem 4.1 and thus R_τ-almost all h_π are strictly singular. Actually (4.1) and a weaker version of (4.2) which just requires that the second moments be bounded imply Theorem 4.1. This leads to the following conjecture which we have been unable to prove.

CONJECTURE: *Theorem 4.2 holds when condition (4.2) is replaced by*

$$\begin{aligned} &\int_{(0,1)} (\ln(y))^2 d\tau(d)(y) < -\ln 2 - a_2, \\ &\int_{(0,1)} (\ln(1-y))^2 d\tau(d)(y) < -\ln 2 - a_2. \end{aligned} \tag{4.2'}$$

On the other hand, Example 4.6 shows that Theorem 4.1 does not imply Theorem 4.2.

Related results for other generating schemes are given by Dubins and Freedman (1967) and Graf, Mauldin, and Williams (1986). The former present conditions under which the distributions they generate are strictly singular almost surely. The latter give conditions so that, for all $x \in [0,1]$, almost all distribution functions have derivative 0 at x. Conversely, Kraft (1964) can be used to state conditions which guarantee that almost all distributions generated by random rescaling are absolutely continuous with respect to Lebesgue measure.

The proof of Theorem 4.1 relies on the following lemmas and uses the scaling map defined in section 2. The lemmas are reformulations and generalizations of Lemmas 5.10, 5.18, and 5.23 of Dubins and Freedman (1967) for our purposes. Also, it will often be convenient to exploit the binary structure of our generating scheme. Let $\{0,1\}^*$ denote the set of all finite sequences of 0's and 1's including the empty sequence $\emptyset = \{0,1\}^0$. Define a map $\beta\colon \{0,1\}^* \to \mathcal{D}$ by $\beta(\emptyset) = \frac{1}{2}$ and, for all other $(b_1, \ldots, b_n) \in \{0,1\}^*$,

$$\beta((b_1, ..., b_n)) = \frac{1}{2} - \sum_{i=1}^{n} \frac{(-1)^{b_i}}{2^{i+1}}.$$

Suppose $b \in \{0,1\}^n$ and $b' \in \{0,1\}^m$, then bb' denotes the element of $\{0,1\}^{n+m}$ whose first n coordinates are b and whose last m coordinates are b'. Take $\emptyset b = b$ and $b\emptyset = b$.

For a transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$, set

$$\tau^* = \tau\beta\colon \{0,1\}^* \to \mathcal{P}([0,1]).$$

The idea is to view τ as a binary tree of probability measures with $\tau(\frac{1}{2}) = \tau^*(\emptyset)$ at the root, $\tau(\frac{1}{4}) = \tau^*(0)$ at the first left node, $\tau(\frac{3}{4}) = \tau^*(1)$ at the first right node, and so on. Set $P_{\tau^*} = \prod_{b \in \{0,1\}^*} \tau^*(b)$. Obviously, any statement about P_{τ^*} and sets in $[0,1]^{\{0,1\}^*}$ will be true for P_τ and the corresponding sets in $[0,1]^{\mathcal{D}}$.

LEMMA 4.3: *Let $K \subset [0,1] - \{\frac{1}{2}\}$ be compact, $t \in [0,1]^{\mathcal{D}}$, n a nonnegative integer, and $x \in [0,1]$ with dyadic expansion $x = .x_1x_2x_3\cdots$. Suppose that for infinitely many nonnegative integers j there exists a $b(j) \in \bigcup_{i=0}^{n}\{0,1\}^i$ such that $t(\beta((x_1, x_2, \ldots, x_j)b(j))) \in K$. Then $\theta(t)$ does not have a finite positive derivative at x.*

Proof: Essentially, the same proof as used in Lemma 5.10 of Dubins and Freedman (1967) works here. ∎

The next few results require some additional notation. For a transition kernel $\tau: \mathcal{D} \to \mathcal{P}([0,1])$ and $b \in \{0,1\}^*$, let $\tau^*[b]: \{0,1\}^* \to \mathcal{P}([0,1])$ be defined by

$$\tau^*[b](b') = \tau^*(bb'),$$

for $b' \in \{0,1\}^*$. Denote $\prod_{b' \in \{0,1\}^*} \tau^*[b](b')$ by $P_{\tau^*[b]}$. Similarly, for $t \in [0,1]^{\mathcal{D}}$ and $b \in \{0,1\}^*$, define $t^*[b] \in [0,1]^{\{0,1\}^*}$ by

$$t^*[b](b') = t^*(bb') = t(\beta(bb')).$$

Let $B \subseteq [0,1]^{\{0,1\}^*}$ and let j and k be nonnegative integers. Define the following sets:

$$\begin{aligned} B^{k,j} = \{t^* \in [0,1]^{\{0,1\}^*}&: \text{for all } (b_1, b_2, \dots) \in \{0,1\} \times \{0,1\} \times \cdots, \\ &\text{there exists } k \text{ or more } n\text{'s such that } t^*[(b_1, \dots, b_n)] \in B \text{ and } n \le j\}, \end{aligned}$$

$$B^k = \bigcup_{j=0}^{\infty} B^{k,j} \quad \text{and} \quad B^\infty = \bigcap_{k=1}^{\infty} B^k.$$

A subset B of $[0,1]^{\{0,1\}^*}$ is **determined by level n** if whenever $t_1^* \in B$ and $t_1^*(b) = t_2^*(b)$ for all $b \in \bigcup_{i=0}^{n} \{0,1\}^i$, then $t_2^* \in B$.

LEMMA 4.4: *Let $\tau: \mathcal{D} \to \mathcal{P}([0,1])$ be a transition kernel, n a nonnegative integer, and $\epsilon > 0$. Suppose B is a Borel subset of $[0,1]^{\{0,1\}^*}$ determined by level n and*

$$2^{n+1}(1 - P_{\tau^*[b]}(B)) < 1 - \epsilon, \tag{4.3}$$

for all $b \in \{0,1\}^*$. Then $P_{\tau^*}(B^\infty) = 1$.

Proof: First, note that it is enough to show that

$$P_{\tau^*[b]}(B^1) = 1 \tag{4.4}$$

for all $b \in \{0,1\}^*$. To see this, suppose (4.4) holds. Then

$$P_{\tau^*}(B^1) = 1.$$

Assume that $P_{\tau^*}(B^i) = 1$, for all $i < k$. Then, for any $\epsilon > 0$, there exists a $j(\epsilon)$, such that

$$P_{\tau^*}(B^{k-1,j(\epsilon)}) > 1 - \epsilon.$$

Let

$$A^j = \{t^* \in [0,1]^{\{0,1\}^*}\colon \text{ for all } (b_1,\ldots,b_{j+1}) \in \{0,1\}^{j+1},\ t^*[(b_1,\ldots,b_{j+1})] \in B^1\}.$$

By (4.4), $P_{\tau^*}(A^j) = 1$, for every j. Hence, using Bonferroni's inequality, for $\epsilon > 0$,

$$P_{\tau^*}(B^k) \geq P_{\tau^*}(B^{k-1,j(\epsilon)} \cap A^{j(\epsilon)}) \geq 1-\epsilon.$$

Since $\epsilon > 0$ was arbitrary, $P_{\tau^*}(B^k) = 1$. Thus, by induction, if (4.4) holds then $P_{\tau^*}(B^\infty) = 1$.

Let $\eta = \inf_{b\in\{0,1\}^*} P_{\tau^*[b]}(B^1)$. If $\eta = 1$, then (4.4), of course, holds. Now let $b \in \{0,1\}^*$. Using the facts that

$$B \subseteq B^1$$

and

$$P_{\tau^*[b]}(B^1|B^c) \geq \prod_{b'\in\{0,1\}^{n+1}} P_{\tau^*[bb']}(B^1),$$

we get

$$\begin{aligned} P_{\tau^*[b]}(B^1) &\geq P_{\tau^*[b]}(B) + (1-P_{\tau^*[b]}(B)) \prod_{b'\in\{0,1\}^{n+1}} P_{\tau^*[bb']}(B^1) \\ &\geq P_{\tau^*[b]}(B) + (1-P_{\tau^*[b]}(B))(\eta)^{2^{n+1}} \end{aligned}$$

Let $b^1, b^2, \ldots$ be a sequence in $\{0,1\}^*$ such that

$$\eta + \frac{1}{m} \geq P_{\tau^*[b^m]}(B^1).$$

Then

$$\eta + \frac{1}{m} \geq P_{\tau^*[b^m]}(B) + (1-P_{\tau^*[b^m]}(B))(\eta)^{2^{n+1}},$$

for all m. By (4.3), there is an $x > 0$ such that $(1-x)2^{n+1} < 1$ and, for all $b \in \{0,1\}^*$, $x \leq P_{\tau^*[b]}(B)$. And so

$$\eta + \frac{1}{m} \geq x + (1-x)(\eta)^{2^{n+1}},$$

for all m. That is,

$$\eta \geq x + (1-x)(\eta)^{2^{n+1}}.$$

A little calculus (for a proof see Lemma 5.15 of Dubins and Freedman (1967)) then shows that $\eta = 1$. ∎

For a subset K of [0,1], denote by $K(n)$ the set

$$K(n) = \left\{\tau^* \in [0,1]^{\{0,1\}^*}\colon \text{ there exists a } b \in \bigcup_{i=0}^{n} \{0,1\}^i \text{ for which } \tau^*(b) \in K\right\}.$$

LEMMA 4.5: *Let $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$ be a transition kernel. Suppose there exists a compact set $K \subseteq [0,1] - \{\frac{1}{2}\}$ and a $\delta > 0$, such that, for all $d \in \mathcal{D}$,*

$$\tau(d)(K) > \delta.$$

Then there exists a nonnegative integer n such that

$$P_{\tau^*}(K(n)^\infty) = 1.$$

Proof: By hypothesis, there exists a nonnegative integer n and $\epsilon > 0$ such that

$$2^{n+1}(1 - \tau^*(b)(K))^{2^{n+1}-1} < 1 - \epsilon,$$

for all $b \in \{0,1\}^*$.

Let $b \in \{0,1\}^*$. Then

$$\begin{aligned} 2^{n+1}(1 - P_{\tau^*[b]}(K(n))) &= 2^{n+1} \prod_{b' \in \bigcup_{i=0}^{n}\{0,1\}^i} \Big(1 - \tau^*(bb')(K)\Big) \\ &\le 2^{n+1}(1 - \tau^*(b\tilde{b})(K))^{2^{n+1}-1} \\ &\le 1 - \epsilon, \end{aligned}$$

where $(1 - \tau^*(b\tilde{b})(K)) = \max_{b' \in \bigcup_{i=0}^{n}\{0,1\}^i}(1 - \tau^*(bb')(K))$. Now apply Lemma 4.4. ∎

Proof of Theorem 4.1: Apply Lemmas 4.3, 4.5, the relationship between P_τ and P_{τ^*}, and the fact that R_τ is induced by P_τ through θ. ∎

Proof of Theorem 4.2: Fix $x \in [0,1]$. Now R_τ-almost all h_π have derivative 0 at x if R_τ-almost all h_π have right-hand derivative 0 at x and R_τ-almost all h_π have left-hand derivative 0 at x. The right-hand case is proved below. The left-hand case can be proved similarly.

As noted in Graf et al. (1986, Theorem 5.20), to show that

$$R_\tau\left(\left\{\pi\colon \lim_{y\downarrow x}\left(\frac{h_\pi(y)-h_\pi(x)}{y-x}\right)=0\right\}\right)=1$$

it is enough to show that, for every $\alpha>0$,

$$R_\tau\left(\bigcap_{k=1}^{\infty}\bigcup_{n\geq k}\left\{\pi\colon \frac{h_\pi(x+2^{-n})-h_\pi(x)}{2^{-n}}\geq\alpha\right\}\right)=0.$$

By the first Borel–Cantelli lemma, this is true if

$$\sum_{n=1}^{\infty}R_\tau\left(\left\{\pi\colon \frac{h_\pi(x+2^{-n})-h_\pi(x)}{2^{-n}}\geq\alpha\right\}\right)<\infty.$$

This holds if for arbitrary $i_n\in\{0,\ldots,2^n-1\}$

$$\sum_{n=1}^{\infty}R_\tau\left(\left\{\pi\colon \frac{h_\pi(\frac{i_n+1}{2^n})-h_\pi(\frac{i_n}{2^n})}{2^{-n}}\geq\frac{\alpha}{2}\right\}\right)<\infty. \tag{4.5}$$

(Again, see Graf et al. (1986, Theorem 5.20).)

Let $\{A_p\}_{d\in\mathcal{D}}$ be a collection of independent random variables defined on a probability space $(\Omega,\mathcal{F},P)$ such that, for $d\in\mathcal{D}$, A_d has distribution $\tau(d)$. Suppose $i\in\{0,\ldots,2^n-1\}$ and $(b_1,\ldots,b_n)\in\{0,1\}^n$ are such that $\sum_{k=1}^n(b_k/2^k)=i/2^n$. Then a straightforward generalization of Lemma 5.21 of Graf et al. (1986) gives

$$\begin{aligned}R_\tau&\left(\left\{\pi\colon \frac{h_\pi(\frac{i_n+1}{2^n})-h_\pi(\frac{i_n}{2^n})}{2^{-n}}\geq\frac{\alpha}{2}\right\}\right)\\&=P\left[\prod_{k=1}^{n}(b_k+(-1)^{b_k}A_{\beta(b_1,\ldots,b_{k-1})})\geq\frac{\alpha}{2^{n+1}}\right],\end{aligned}$$

with the convention that, when $k=1$, $A_{\beta(b_1,\ldots,b_{k-1})}=A_{\frac{1}{2}}$.

From condition (4.1) we get

$$\begin{aligned}
&P\left[\prod_{k=1}^{n}(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})\geq\frac{\alpha}{2^{n+1}}\right]\\
&=P\left[\sum_{k=1}^{n}\ln(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})\geq\ln(\frac{\alpha}{2})-n\ln 2\right]\\
&=P\left[\sum_{k=1}^{n}\ln(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})-u_{(b_1,\dots,b_{k-1})}\right.\\
&\qquad\left.\geq\ln(\frac{\alpha}{2})-\sum_{k=1}^{n}(-\ln 2-u_{(b_1,\dots,b_{k-1})})\right]\\
&\leq P\left[\sum_{k=1}^{n}\ln(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})-u_{(b_1,\dots,b_{k-1})}\geq\ln(\frac{\alpha}{2})+na_1\right],
\end{aligned}$$

where $u_{(b_1,\dots,b_{k-1})}=E[\ln(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})]$.

Let $S_n=\sum_{k=1}^{n}\ln(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})-u_{(b_1,\dots,b_{k-1})}$. Then, for n large enough that $\ln(\alpha/2)+na_1>0$, Markov's inequality gives

$$\begin{aligned}
&P\left[\sum_{k=1}^{n}\ln(b_k+(-1)^{b_k}A_{\beta(b_1,\dots,b_{k-1})})-u_{(b_1,\dots,b_{k-1})}\geq\ln(\alpha/2)+na_1\right]\\
&\qquad\leq\frac{E[S_n^4]}{(\ln(\alpha/2)+na_1)^4}.
\end{aligned}$$

Moreover, by condition (4.2) and the independence of the A_p's, there exists a constant M which does not depend on n such that

$$E[S_n^4]\leq Mn^2.$$

Thus (4.5) is true. ∎

Note that Theorem 4.2 is still true if (4.1) is weakened somewhat to condition (4.6), given below. For $b\in\{0,1\}^*$, let

$$u_{b,0}=\int_{(0,1)}\ln(x)d\tau(\beta(b))(x)$$

and

$$u_{b,1}=\int_{(0,1)}\ln(1-x)d\tau(\beta(b))(x).$$

CONDITION 4.6: *There exists a constant $a > 0$ and an integer N such that, for every $n > N$ and $(b_1, \ldots, b_n) \in \{0,1\}^n$,*

$$-\ln 2 - \frac{1}{n}\sum_{k=1}^{n} u_{(b_1,\ldots,b_{k-1}),b_k} > a,$$

with the convention that, when $k = 1$, $(b_1, \ldots, b_{k-1}) = \emptyset$.

Example 4.6 shows that (4.1) alone is not sufficient to guarantee the conclusion of Theorem 4.2. In particular, it shows that it is not sufficient to bound the variances (or fourth moments) of the $\ln(A_d)'s$ and the $\ln(1-A_d)'s$ (A_d is as given in the proof of Theorem 4.2) along each branch of the binary tree associated with τ. As mentioned, we believe that if the variances are uniformly bounded then the conclusion of the theorem holds. Notice that the transition kernel in the example satisfies the conditions of Theorem 4.1 — so almost all distribution functions are strictly singular. Moreover, by Corollary 3.4, almost all distribution functions are continuous.

Example 4.6: For $c = 2, 3, \ldots$, let

$$a_c = \frac{(c - \frac{3}{2})\ln 2}{c\ln(2) + \ln\frac{3}{4}}$$

and

$$\mu_c = (1 - a_c)\delta_{\{1/2^c\}} + a_c\delta_{\{3/4\}}.$$

Then there exists a positive integer C such that, for $c \geq C$,

$$\int_{(0,1)} \ln(x)d\mu_c(x) = -\frac{3}{2}\ln 2$$

and

$$\int_{(0,1)} \ln(1 - x)d\mu_c(x) \leq -\frac{3}{2}\ln 2.$$

Also, $\lim_{c\to\infty} a_c = 1$.

Select a subsequence of the $a_c's$, $a_{c_1}, a_{c_2}, \ldots$, such that $\prod_{k=1}^{\infty} a_{c_k} > z > 0$ and, for each k, $c_k > C$. Define n_1 and m_1 by

$$n_1 = 1 \quad \text{and} \quad m_1 = \lfloor\frac{n_1\ln 2}{\ln\frac{3}{2}}\rfloor,$$

where $\lfloor y \rfloor$ is the least integer greater than y. (Note that $m_1 = 2$.) For $i = 2, 3, \ldots,$ set

$$n_i = n_{i-1} + m_{i-1} + 1 \quad \text{and} \quad m_i = \lfloor \frac{n_i \ln 2}{\ln \frac{3}{2}} \rfloor.$$

Let $x \in (0,1)$ be that number with binary expansion

$$x = .b_1 b_2 b_3 \cdots = .0110\underbrace{1\ldots1}_{m_2\ 1's}01\ldots101\ldots10\underbrace{1\ldots1}_{m_i\ 1's}\ldots$$

Define a map $\tau^*\colon \{0,1\}^* \to \mathcal{P}([0,1])$ by

$$\begin{aligned} \tau^*(1) &= \mu_{c_1}, \\ \tau^*(1,0) &= \mu_{c_2}, \end{aligned}$$

for $k = 2, 3, \ldots,$

$$\begin{aligned} \tau^*(b_1, \ldots, b_{n_k-1}, 1) &= \mu_{c_1}, \\ \tau^*(b_1, \ldots, b_{n_k-1}, 1, 0) &= \mu_{c_2}, \\ &\vdots \\ \tau^*(b_1, \ldots, b_{n_k-1}, 1, \underbrace{0, \ldots, 0}_{m_k-1\ 0\text{'s}}) &= \mu_{c_{m_k}} \end{aligned}$$

and for all other $b \in \{0,1\}^*$

$$\tau^*(b) = \mu_2.$$

Now define a transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$ by $\tau = \tau^* \beta^{-1}$.

Letting $x_n = \sum_{i=1}^{n} b_i/2^i$,

$$\begin{aligned} R_\tau(\{\pi\colon h'_\pi(x) = 0\}) &\le R_\tau\left(\left\{\pi\colon \lim_{n\to\infty} \frac{h_\pi(x_n + 2^{-n+1}) - h_\pi(x_n)}{2^{-n+1}} = 0\right\}\right) \\ &= R_\tau\left(\bigcap_{\gamma>0} \bigcup_{j=1}^{\infty} \bigcap_{k\ge j} \left\{\pi\colon \frac{h_\pi(x_k + 2^{-k+1}) - h_\pi(x_k)}{2^{-k+1}} < \gamma\right\}\right). \end{aligned}$$

Suppose $\gamma = \frac{1}{2}$, then

$$R_\tau\left(\bigcup_{j=1}^{\infty}\bigcap_{k\geq j}\left\{\pi\colon \frac{h_\pi(x_k+2^{-k+1})-h_\pi(x_k)}{2^{-k+1}}<\gamma\right\}\right)$$

$$=1-R_\tau\left(\bigcap_{j=1}^{\infty}\bigcup_{k\geq j}\left\{\pi\colon \frac{h_\pi(x_k+2^{-k+1})-h_\pi(x_k)}{2^{-k+1}}\geq\frac{1}{2}\right\}\right)$$

$$\leq 1-R_\tau\left(\bigcap_{j=1}^{\infty}\bigcup_{k\geq j}\left\{\pi\colon \frac{h_\pi(x_k+2^{-k+1})-h_\pi(x_k+2^{-k})}{2^{-k}}\geq 1\right\}\right)$$

$$=1-R_\tau\left(\bigcap_{j=1}^{\infty}\bigcup_{k\geq j}\left\{\pi\colon h_\pi(x_k+2^{-k+1})-h_\pi(x_k+2^{-k})\geq\frac{1}{2^k}\right\}\right)$$

$$\leq 1-R_\tau\left(\bigcap_{j=2}^{\infty}\bigcup_{k\geq j}\left\{\pi\colon h_\pi(x_{n_k-1}+2^{-n_k+2})-h_\pi(x_{n_k-1}+2^{-n_k+1})\geq\frac{1}{2^{n_k-1}}\right\}\right)$$

$$=1-P[\bigcap_{j=2}^{\infty}\bigcup_{k\geq j}B_k],$$

where $B_2=[(1-A_{\beta(\emptyset)})A_{\beta(1)}A_{\beta(1,0)}\geq\frac{1}{8}]$ and, for $k>2$,

$$B_k=\left[\left(\prod_{i=1}^{n_{k-1}-1}(b_i+(-1)^{b_i}A_{\beta(b_1,\ldots,b_{i-1})}\right)(1-A_{\beta(b_1,\ldots,b_{n_{k-1}-1})})\right.$$
$$A_{\beta(b_1,\ldots,b_{n_{k-1}-1},1)}A_{\beta(b_1,\ldots,b_{n_{k-1}-1},1,0)},\ldots,$$
$$\left.A_{\beta(b_1,\ldots,b_{n_{k-1}-1},1,\underbrace{0,\ldots,0}_{m_{k-1}-1\ 0\text{'s}})}\geq\frac{1}{2^{n_k-1}}\right]$$

(notation as in the proof of Theorem 4.2).

By the construction of τ

$$E_2=[A_{\beta(1)}A_{\beta(1,0)}\geq\frac{1}{2}]\subseteq B_2,$$

and for $k>2$

$$E_k=\left[A_{\beta(b_1,\ldots,b_{n_{k-1}-1},1)}A_{\beta(b_1,\ldots,b_{n_{k-1}-1},1,0)},\ldots,\right.$$
$$\left.A_{\beta(b_1,\ldots,b_{n_{k-1}-1},1,\underbrace{0,\ldots,0}_{m_{k-1}-1\ 0'\text{s}})}\geq\frac{2^{n_{k-1}}}{2^{m_{k-1}}}\right]\subseteq B_k.$$

Also, the $E_k's$ are independent with $P[E_k] \geq z$, for all $k \geq 2$. Thus, by the second Borel–Cantelli lemma

$$1 - P[\bigcap_{j=2}^{\infty} \bigcup_{k \geq j} B_k] \leq 1 - P[\bigcap_{j=2}^{\infty} \bigcup_{k \geq j} E_k] = 0.$$

Therefore, $R_\tau(\{\pi: h'_\pi(x) = 0\}) = 0$. ∎

5. Hausdorff dimension of the generated probability measures

Lastly, we investigate the (Hausdorff) dimension of the probability measures generated by random rescaling. We establish bounds between which the dimension of almost all measures lie. Note in the special case that $\tau(d) = \mu$, for all $d \in \mathcal{D}$, the results below correspond to a special case of Kinney and Pitcher (1964). Moreover, Theorem 1 of Kinney and Pitcher (1964) is related to Lemma 5.1 below. However, there appears to be an error in their proof of the theorem. The hypotheses they give in their theorem are not quite what is needed for their conclusion. What is needed is a condition analogous to that given in Lemma 5.1. In the remainder of this section, assume that all the transition kernels considered are centered. So, R_τ-almost all distribution functions are continuous.

Recall that the **Hausdorff dimension of a probability measure** $\pi \in \mathcal{P}([0,1])$ is

$$\dim_{\mathcal{H}}(\pi) = \min\{\dim_{\mathcal{H}}(A): A \subseteq [0,1] \text{ and } \pi(A) = 1\},$$

where $\dim_{\mathcal{H}}(A)$ is the Hausdorff dimension of A. That is, for $\delta > 0$,

$$\mathcal{H}_\delta^\alpha(A) = \min\left\{\sum_{G \in \mathcal{G}} |G|^\alpha: \mathcal{G} \text{ is a } \delta\text{-mesh cover of } A\right\}$$

and the **α-dimensional measure** of A is

$$\mathcal{H}^\alpha(A) = \lim_{\delta \to 0} \mathcal{H}_\delta^\alpha(A).$$

The **Hausdorff dimension** of A is

$$\dim_{\mathcal{H}}(A) = \inf\{\alpha: \mathcal{H}^\alpha(A) = 0\} = \sup\{\alpha: \mathcal{H}^\alpha(A) = \infty\}.$$

For $x \in [0,1)$ and $n \geq 0$, define $\alpha_n(x)$ such that $\alpha_n(x) = i/2^n \leq x < (i+1)/2^n$. And if $.b_1 b_2 \ldots b_n$ is the nth-order dyadic expansion of $i/2^n$, let $b_0(i/2^n) = \emptyset$ and, for $1 \leq k \leq n$, let $b_k(i/2^n) = (b_1, \ldots, b_k)$.

For a transition kernel $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$ and $b \in \{0,1\}^*$, set

$$\gamma_b = \int_{[0,1]} y\ln(y) + (1-y)\ln(1-y)\; d\tau^*(b)(y).$$

And, for $n \geq 1$, let

$$\begin{aligned}\sigma_n^2 &= \sup_{b\in\{0,1\}^{n-1}} \left\{ \int y(\ln(y)-\gamma_b)^2 + (1-y)(\ln(1-y)-\gamma_b)^2\; d\tau^*(b)(y) \right\} \\ &= \sup_{b\in\{0,1\}^{n-1}} \left\{ \int y\ln^2(y) + (1-y)\ln^2(1-y)\; d\tau^*(b)(y) - \gamma_b^2 \right\}.\end{aligned}$$

As above, we often identify a probability measure with its distribution function. Denote the set of distribution functions on [0,1] by H.

LEMMA 5.1: *Let $\tau\colon \mathcal{D} \to \mathcal{P}([0,1])$ be a centered transition kernel. Suppose*

(i) $\int_{[0,1]} \left|y\ln(y) + (1-y)\ln(1-y)\right| d\tau(d)(y) < \infty$, *for all $d \in \mathcal{D}$,*

(ii) $\sum_{n=1} \sigma_n^2/n^2 < \infty$, *and*

(iii) $\gamma_x = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \gamma_{b_j(\alpha_n(x))}$ *exists, for all $x \in [0,1]$.*

Then

$$\lim_{n\to\infty} \frac{1}{n}\left(\ln\left(h_\pi\left(\alpha_n(x)+\frac{1}{2^n}\right) - h_\pi(\alpha_n(x))\right)\right) = \gamma_x$$

for h_π-almost all $x \in [0,1]$ and R_τ-almost all $h_\pi \in H$.

Proof: For $n \geq 1$, define $f_n\colon [0,1]\times H \to \mathbb{R}$ by

$$f_n(x,h) = \ln\left(\frac{h(\alpha_n(x)+\frac{1}{2^n}) - h(\alpha_n(x))}{h(\alpha_{n-1}(x)+\frac{1}{2^{n-1}}) - h(\alpha_{n-1}(x))}\right) - \gamma_{b_{n-1}(\alpha_n(x))}.$$

Let $\mathcal{F}_n$ be the σ-algebra generated by $\{[i/2^n,(i+1)/2^n]\times H\}_{0\leq i\leq 2^n-1}$ and define a measure Ψ on $[0,1]\times H$ by

$$\int f(x,h_\pi)d\Psi = \int_H \int_{[0,1]} f(x,h)dh_\pi(x)dR_\tau(\pi).$$

Then, for any $0 \leq i \leq 2^{n-1}-1$ and $n \geq 1$,

$$\begin{aligned}&E_\Psi\left[I_{[i/2^{n-1},(i+1)/2^{n-1}]\times H}\; f_n(x,h_\pi)\right] \\ &= \int_H \int_{[1/2^{n-1},(i+1)/2^{n-1}]} \ln\left(\frac{h_\pi(\alpha_n(x)+1/2^n) - h_\pi(\alpha_n(x))}{h_\pi(\alpha_{n-1}(x)+1/2^{n-1}) - h_\pi(\alpha_{n-1}(x))}\right) \\ &\qquad\qquad - \gamma_{b_{n-1}(\alpha_n(x))}\; dh_\pi(x)dR_\tau(\pi) \\ &= \int_H \left(h_\pi\left(\frac{i+1}{2^{n-1}}\right) - h_\pi\left(\frac{i}{2^{n-1}}\right)\right)\left(\int_0^1 y\ln(y)+(1-y)\ln(1-y)\right. \\ &\qquad\qquad \left. d\tau^*(b_{n-1}(i/2^{n-1})) - \gamma_{b_{n-1}(i/2^{n-1})}\right) dR_\tau(\pi) \\ &= 0.\end{aligned}$$

Thus, $E_\Psi[f_n|\mathcal{F}_{n-1}] = 0$. Similarly it can be checked that $E_\Psi[f_n^2] \leq \sigma_n^2$. It now follows that $S_n = \sum_{j=1}^n f_j/j$ is an $\{\mathcal{F}_n\}$-martingale such that $E_\Psi[S_n^2] \leq \sum_{j=1}^n \sigma_j^2/j^2 < \infty$. Apply the martingale convergence theorem (Dellacherie and Meyer (1982, Theorem V.30)) to conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n f_j = \lim_{n\to\infty}\left(S_n - \frac{1}{n}\sum_{j=1}^{n-1} S_j\right) = 0. \qquad \blacksquare$$

We can now give bounds on the dimensions of almost all measures generated by random rescaling. Our proof may be viewed as a "Frostman" type lemma. (Note that, where it exists, γ_x is negative.)

THEOREM 5.2: *Suppose that the hypotheses of Lemma* 5.1 *hold. Let*

$$\gamma_1 = \inf\{\gamma_x \colon x \in [0,1]\} \quad \textit{and} \quad \gamma_2 = \sup\{\gamma_x \colon x \in [0,1]\}.$$

Then, for R_τ-almost all π,

$$\frac{-\gamma_2}{\ln 2} \leq \dim_{\mathcal{H}}(\pi) \leq \frac{-\gamma_1}{\ln 2}.$$

Proof: Suppose $\pi \in \mathcal{P}([0,1])$ is continuous and that for h_π-almost all $x \in [0,1]$

$$\lim_{n\to\infty} \frac{1}{n}\left(\ln\left(h_\pi(\alpha_n(x) + \frac{1}{2^n}) - h_\pi(\alpha_n(x))\right)\right) = \gamma_x. \tag{5.1}$$

By Lemma 5.12, the set of such h_π has R_τ measure 1. Let A be a set of h_π measure 1 such that, for all $x \in A$, (5.1) holds. Fix $\epsilon, \delta > 0$. We will show that $\mathcal{H}_\delta^\eta(A) \leq 1$, where $\eta = -\gamma_1/\ln 2 + \epsilon$.

For each $x \in A$, let $n(x)$ be the smallest n such that

$$\frac{1}{2^n} < \delta \quad \text{and} \quad h_\pi(\alpha_n(x) + \frac{1}{2^n}) - h_\pi(\alpha_n(x)) > \left(\frac{1}{2^n}\right)^{-\gamma_1/\ln 2 + \epsilon}$$

Notice that, for $x_1, x_2 \in [\alpha_{n(x)}(x), \alpha_{n(x)}(x) + 1/2^{n(x)}] \cap A$, $n(x_1) = n(x_2)$. Thus,

$$\left\{\left[\alpha_{n(x)}(x), \alpha_{n(x)}(x) + 1/2^{n(x)}\right]\right\}_{x\in A}$$

is a disjoint (except for endpoints) countable cover of A. Therefore,

$$\begin{aligned}
1 &= \pi(A) \\
&= \pi\left(\bigcup\left[\alpha_{n(x)}(x), \alpha_{n(x)}(x) + \frac{1}{2^{n(x)}}\right]\right) \\
&= \sum\left(h_\pi(\alpha_{n(x)}(x) + \frac{1}{2^{n(x)}}) - h_\pi\left(\alpha_{n(x)}(x)\right)\right) \\
&\geq \sum\left(\frac{1}{2^{n(x)}}\right)^{-\gamma_1/\ln 2 + \epsilon}
\end{aligned}$$

Since δ and ϵ were arbitrary, $\mathcal{H}(A) \leq -\gamma_1/\ln 2$; and so,

$$\dim_{\mathcal{H}}(\pi) \leq \frac{-\gamma_1}{\ln 2}.$$

Again fix $\epsilon > 0$ and set

$$K(\epsilon) = \{x \in [0,1]\colon h_\pi(\alpha_n(x) + 1/2^n) - h_\pi(\alpha_n(x)) > \left(\frac{1}{2^n}\right)^{-\gamma_2/\ln 2 - \epsilon} \text{ for infinitely many } n\}.$$

For $\delta > 0$ and $x \in [0,1]$, let $m(x)$ be the smallest n (for x where such an n exists) such that

$$\frac{1}{2^n} < \delta \quad \text{and} \quad h_\pi\Big(\alpha_n(x) + \frac{1}{2^n}\Big) - h_\pi(\alpha_n(x)) > \left(\frac{1}{2^n}\right)^{-\gamma_2/\ln 2 - \epsilon}$$

Set

$$K(\epsilon, \delta) = \left\{\left[\alpha_{m(x)}(x), \alpha_{m(x)}(x) + \frac{1}{2^{m(x)}}\right]\right\}_{x \in [0,1]}$$

Then $K(\epsilon,\delta)$ covers $K(\epsilon)$ and $\bigcap_{\delta\to 0} K(\epsilon,\delta)$ differs from $K(\epsilon)$ by at most a countable set. Therefore, since π is continuous,

$$0 = \pi(K(\epsilon)) = \pi\Big(\bigcap_{\delta\to 0} K(\epsilon,\delta)\Big) = \lim_{\delta\to 0} \pi(K(\epsilon,\delta)).$$

Suppose $B \subseteq [0,1]$ such that $\pi(B) = 1$. Then there exists $f(\epsilon,\delta)$ such that

$$\pi\Big((I \cap A \cap B) - K(\epsilon,\delta)\Big) < |I|^{(-\gamma_2/\ln 2) - f(\epsilon,\delta)},$$

for any interval I for which $|I| < \delta$. Moreover, $f(\epsilon,\delta)$ can be chosen so that $f(\epsilon,\delta)$ decreases (to ϵ) as δ decreases and $f(\epsilon,\delta) \to 0$ as ϵ and $\delta \to 0$.

Now fix $\hat{\delta} > 0$. We will show that $\mathcal{H}^\kappa_{\hat{\delta}}(B) \geq \frac{1}{2}$, where $\kappa = -\gamma_2/\ln 2 - f(\epsilon,\hat{\delta})$. Let $C = \{I_n\}$ be a set of intervals covering $(A \cap B) - K(\epsilon,\hat{\delta})$ with $|I_n| < \delta < \hat{\delta}$ and $\pi(K(\epsilon,\hat{\delta})) < \frac{1}{2}$. Then

$$\begin{aligned}\frac{1}{2} \leq \pi((A \cap B) - K(\epsilon,\hat{\delta})) &\leq \sum_n \pi((I_n \cap A \cap B) - K(\epsilon,\delta)) \\ &\leq \sum_n |I_n|^{(-\gamma_2/\ln 2) - f(\epsilon,\hat{\delta})}.\end{aligned}$$

Hence

$$\frac{1}{2} \leq \mathcal{H}^{(-\gamma 2/\ln 2) - f(\epsilon,\hat{\delta})}((A \cap B) - K(\epsilon,\hat{\delta})) \leq \mathcal{H}^{(-\gamma^2/\ln 2) - f(\epsilon,\hat{\delta})}(B).$$

Since $f(\epsilon,\delta)$ can be arbitrarily small, it follows that $\frac{-\gamma_2}{\ln 2} \leq \mathcal{H}(B)$; and so,

$$\frac{-\gamma_2}{\ln 2} \leq \dim_{\mathcal{H}}(\pi). \quad \blacksquare$$

References

[1] D. J. Aldous, *Exchangeability and related topics*, Ecole d'Ete de Probabilite de Saint-Flour XIII, Lecture Notes in Math **1117** (1983), 1–197.

[2] C. Dellacherie and P. Meyer, *Probabilities and Potential B*, North-Holland, Amsterdam, 1982.

[3] K. Doksum, *Tailfree and neutral random probabilities and their posterior distributions*, The Annals of Probability **2** (1974), 183–201.

[4] L. E. Dubins and D. A. Freedman, *Random distribution functions*, Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability **2** (1967), 183–214.

[5] T. S. Ferguson, *A Bayesian analysis of some nonparametric problems*, The Annals of Statistics **1** (1973), 209–230.

[6] T. S. Ferguson, *Prior distributions on spaces of probability measures*, The Annals of Statistics **2** (1974), 615–629.

[7] S. Graf, R. D. Mauldin and S. C. Williams, *Random homeomorphisms*, Advances in Mathematics **60** (1986), 239–359.

[8] S. Graf, E. Novak, and A. Papageorgiou, *Bisection is not optimal on the average*, Numerische Mathematik **55** (1989), 481–491.

[9] J. R. Kinney and T. S. Pitcher, *The dimension of the support of a random distribution function*, Bulletin of the American Mathematical Society **70** (1964), 161–164.

[10] C. H. Kraft, *A class of distribution function processes which have derivatives*, Journal of Applied Probability **1** (1964), 385–388.

[11] R. D. Mauldin, W. D. Sudderth and S. C. Williams, *Polya trees and random distributions*, The Annals of Statistics **20** (1992), 1203–1221.

[12] E. Novak, *Stochastic properties of quadrature formulas*, Numerische Mathematik **53** (1988), 609–620.

[13] E. Novak, *Average-case results in zero finding*, Journal of Complexity **5** (1989), 489–501.

[14] E. Novak and K. Ritter, *Some complexity results for zero finding for univariate functions*, preprint, 1992.

[15] K. Ritter, *Average errors for zero finding: lower bounds for smooth or monotone functions*, University of Kentucky Technical Report No. 209-92, 1992.

[16] S. M. Ulam, *Transformation, iterations, and mixing flows*, in *Dynamical Systems II* (A. R. Bednarek and L. Cesari, eds.), Academic Press, New York, 1982, pp. 419–426