Statistica Sinica **16**(2006), 77-91

MEETING HAUSDORFF IN MONTE CARLO: A SURPRISING TOUR WITH ANTIHYPE FRACTALS

Radu V. Craiu and Xiao-Li Meng

University of Toronto and Harvard University

Abstract: To many statistical researchers, fractals are aesthetically pleasing mathematical objects or ingredients of complex theoretical studies. This article documents an exception: during recent research on improving effectiveness of Markov chain Monte Carlo (MCMC), we unexpectedly encountered a class of intriguing fractals in the simple context of generating negatively correlated random variates that achieve extreme antithesis. This class of *antihype fractals* enticed us to tour the world of fractals, because it has intrinsic connections with classical fractals such as Koch's snowflake and it illustrates theoretical concepts such as Hausdorff dimension in a very intuitive way. It also provides a practical example where a sequence of uniform variables converges exponentially in the Kolmogorov-Smirnov distance, yet fails to converge in other common distances, including total variation distance and Hellinger distance. We also show that this non-convergence result actually holds for any sequence of (proper) uniform distributions on supports formed by the generating process of a self-similar fractal. These negative results remind us that the choice of metrics, e.g., for diagnosing convergence of MCMC algorithms, do matter sometimes in practice.

Key words and phrases: Antithetic variates, extreme antithesis, fractals, Hausdorff dimension, Koch's curve, Latin hypercube sampling, Markov chain Monte Carlo, self-similar fractals.

1. Getting Started - Fractals

The eye-pleasers of this tour are the fractals. Although it is difficult to define precisely what constitutes a fractal, one may consider that any set with a high degree of irregularity present at any scale would fit into this category. Benoit Mandelbrot, the French mathematician who brought fractals into mainstream science, emphasizes that such irregular geometric abstractions fit the physical world better: "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line" (Mandelbrot (1982)).

Besides their intricate and intriguing appearance, fractals also furthered the exploration of mathematical concepts. For instance, consider the Koch's snowflake as discussed in Edgar (1990). Take an equilateral triangle and divide

each of its sides into three equal parts. Replace the middle segment by the other two sides of an equilateral triangle constructed on the middle segment. Continue the same process on each of the remaining segments in the newly obtained figure. The first two steps of the iterative process are sketched in Figure 1. The limit, shown in Figure 2, is the remarkable Koch's curve. It surrounds a finite area, yet it has infinite length. It is continuous, yet nowhere differentiable. It is also self-similar: under magnification, arbitrarily small regions look exactly the same as the whole.

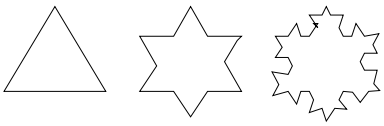

Figure 1. Koch's construction.

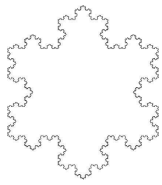

Figure 2. Koch's snowflake.

Soon after its discovery in 1906, there was a general recognition of the inadequacy of classical tools to deal with objects as irregular as the Koch's snowflake. Hausdorff (1919) extended the work of his contemporaries, Borel and Frechet, by introducing the Hausdorff measure and dimension. In particular, it allows a non-integer dimension which is, as we demonstrate, a very intuitive generalization of the Euclidian dimension that we normally encounter in classical geometry.

To respect the tradition of good tourism, our tour combines information and entertainment, the past and the future. We first visit the construction of *antihype fractals*, unearthed while generating *anti*thetic random variates via Latin *hype*rcube sampling. We then enter the world of Hausdorff dimension, starting from its abstract land, with a detour to a visually attractive connection between Koch's snowflake and an antihype fractal. Leaving Hausdorff, we visit a few other masters of abstraction, Kolmogorov, Smirnov, and Hellinger — if you wonder why, please follow the guide!! (And please do pick up a couple of tour souvenirs we prepared along the tour.)

2. Antihype Fractals

2.1. The Monte Carlo application

A well-known Monte Carlo technique is to use variates with negative correlations. Let $\{X_1, \ldots, X_k\}$ be k exchangeable draws with $\mathrm{Corr}(X_i, X_j) = \rho_k$ and $\mathrm{Var}(X_i) = \sigma^2$. Then, it is easy to show that

$$\mathrm{Var}(\bar{X}_k) \equiv \mathrm{Var}\Big(\frac{1}{k}\sum_{i=1}^{k} X_i\Big) = \frac{\sigma^2}{k}[1 + (k-1)\rho_k].$$

Consequently, if $\rho_k < 0$, the right-hand side is smaller than σ^2/k, the variance of $\bar{X}_k$ when $\{X_1, \ldots, X_k\}$ are uncorrelated. In particular if, and only if, $\rho_k = -(k-1)^{-1}$, then $\mathrm{Var}(\bar{X}_k) = 0$. In general, for an arbitrary given marginal distribution of X_i and integer k, the smallest possible ρ_k always exists but may not reach $-(k-1)^{-1}$ (e.g., for exponential distributions, as reported in Moran (1967)). A vector $\{X_1, \ldots, X_k\}$ achieving the minimum possible ρ_k is said to have the *extreme antithesis (EA)* property (Craiu and Meng (2005)). Since uniform random variables are the most basic building blocks for simulation, we focus on $X = U \sim \mathrm{Uniform}[0,1]$ in this paper.

Suppose that we want to sample $(U^{(1)}, U^{(2)})$ in $[0,1]^2$ with $\rho_2 = -1$. Instead of rushing to the obvious solution $U^{(2)} = 1 - U^{(1)}$, one could envision the following approach. At draw $t = 1, 2, \ldots$, we sample randomly from the finite set $\{0,1\}$, *without replacement*, and use the draws as the dyadic coefficients. That is, we sample $a_t \in \{0,1\}$ independently, set $b_t = 1 - a_t$, then let

$$\begin{aligned} \tilde{V}_t^{(1)} &= \frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_t}{2^t} \\ \tilde{V}_t^{(2)} &= \frac{b_1}{2} + \frac{b_2}{2^2} + \cdots + \frac{b_t}{2^t}. \end{aligned} \tag{2.1}$$

Clearly, as t increases, the pair $(\tilde{V}_t^{(1)}, \tilde{V}_t^{(2)})$ converges to the well-known solution $(U, 1-U)$ because $\tilde{V}_t^{(1)} + \tilde{V}_t^{(2)} = 1 - 2^{-t}$. This simple construction has a defect: for any *finite* t, the marginal distribution of $\tilde{V}_t^{(i)}$ is not exactly $\mathrm{Uniform}(0,1)$. However, we can easily correct for this because

$$\tilde{U}_t^{(i)} = V_t^{(i)} + \frac{U_0^{(i)}}{2^t} \sim \mathrm{Uniform}(0,1), \quad i = 1, 2, \tag{2.2}$$

as long as $U_0^{(i)} \sim \mathrm{Uniform}(0,1)$ and it is independent of $V_t^{(i)}$.

2.2. Iterative Latin hypercube sampling

The construction (2.1) is suggestive for going beyond $k = 2$: replace the dyadic expansion in (2.1) with the k-based expansion. Specifically, Craiu and Meng (2005) constructed the following iterative scheme, to which (2.2) is the "time backward dual sequence" (when $k = 2$; see Section 5.1).

1. Set $\mathcal{U}_0^k = (U_0^{(1)}, \ldots, U_0^{(k)})^\top$, where $\{U_0^{(i)}\}_{1 \le i \le k}$ are i.i.d. $\mathrm{Uniform}(0,1)$.
2. For $t = 1, \ldots$, draw a random permutation of $\{0, 1, \ldots, k-1\}$, $\mathcal{K}_t = (\sigma_t(0), \ldots, \sigma_t(k-1))^\top$, which is independent of all previous draws. We then define $\mathcal{U}_t^k = (U_t^{(1)}, \ldots, U_t^{(k)})^\top$ via

$$\mathcal{U}_{t+1}^k = \frac{1}{k}(\mathcal{K}_t + \mathcal{U}_t^k). \tag{2.3}$$

This *iterative Latin hypercube sampling* (ILHS) scheme is an iterative version of the well-known Latin hypercube sampling (McKay, Beckman and Conover (1979), Stein (1987), Owen (1992) and Iman (1999)). A more intuitive perspective on ILHS can be visualized in the case $k = 3$. Figure 3 plots the supports of $\mathcal{U}_t^3$ for $t = 1, 2$. Starting from the unit cube $(0, 1)^3$, the tth iteration of ILHS divides the unit cube into 3^{3t} sub-cubes, each with volume 3^{-3t}, and eliminates the "most positive" three sub-sub-cubes from each of the sub-cubes that *survived* after the previous iteration/elimination. For $t = 1$, this is seen in the left plot of Figure 3. Evidently, the three excluded sub-cubes induce the highest positive correlation between any pair of the components of $\mathcal{U}_t^3$. The right plot repeats the same process within each of the $3^3 - 3$ sub-cubes that remain after the first elimination; the total number of sub-cubes of volume 3^{-3t} eliminated at t-th ($t > 0$) iteration is $3(3^3 - 3)^{t-1} = 8^{t-1}3^t$.

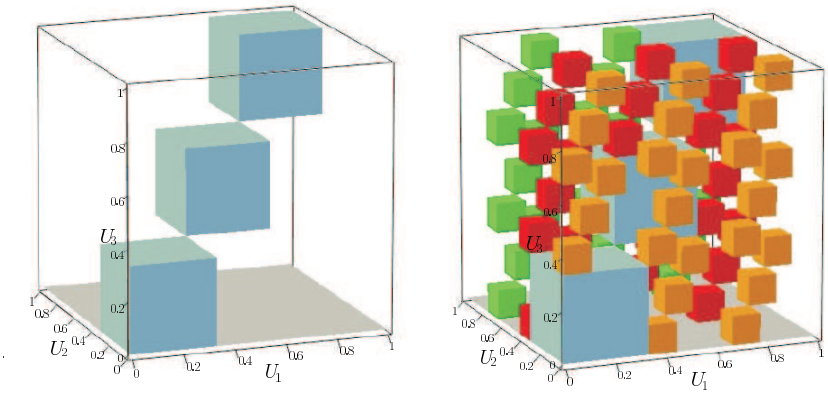

Figure 3. Evolution of the support of $\mathcal{U}_t^3$ for $t = 1, 2$. The regions eliminated at each step are shown in colors.

As proved in Craiu and Meng (2005), as t increases, the vector $\mathcal{U}_t^k$ gets closer and closer to achieving EA, and the limit $\mathcal{U}_\infty^k$ achieves EA exactly. Furthermore, the support of $\mathcal{U}_t^k$ shrinks as t increases, as seen in Figure 3. Indeed, each iteration of ILHS eliminates $p = k^{-(k-1)}$ portion of the volume left from the previous iteration. Summing up all these volumes from $t = 1$ to $t = \infty$, that is, $\sum_{t=1}^{\infty} p(1-p)^{t-1} = 1$, we see that the volume of the support of the limiting variable $\mathcal{U}_\infty^k$, is *zero* in terms of the usual Lebesgue measure. Somewhat remarkably, the marginal uniformity of each $U_\infty^{(j)}$ is preserved. This Lebesgue null set is our antihype fractal and is denoted by $\mathcal{A}^k$.

The properties of the $\mathcal{A}^k$ are important to the study of efficient coupling in the context of Markov chain Monte Carlo; see Section 6. Chatterjee and Yilmaz (1992) is a good place to sample other connections between fractals and

statistics. The bulk of interaction lies in the area of dynamical systems for number generation (e.g., Palmore (1994) and Gerow and Holbrook (1996)), noise reduction for reconstruction of time series data (e.g., Kostelich and Yorke (1988)) and fractal interpolation (Barnsley (1988)). The antihype fractal belongs to the first category.

3. Fractal Dimensions

3.1. Defining the Hausdorff dimension

Sets of elementary geometry have an associated dimension, e.g., points have dimension zero, curves one, surfaces two, and so on. Once we leave elementary geometry we encounter sets that do not fall into any of the above categories, although they "live" in the Euclidean space. For such sets, such as Koch's curve or antihype fractals, one needs to generalize the usual notion of dimension. Among the several "fractal dimensions" in use, the definition of Hausdorff is the oldest and, probably, the most important. This is also the dimension singled out by Mandelbrot as the most effective in establishing interesting geometric properties, although it is not always easy to compute. The formal definition of the Hausdorff dimension of a fractal set, $\dim_H(\mathcal{F})$, is given below.

Definition 1. Suppose that $\mathcal{F}$ is a subset in R^d.

1. If $\{U_i\}$ is a countable collection of sets of diameter at most δ, whose union contains $\mathcal{F}$, we say that $\{U_i\}$ is a δ-cover of $\mathcal{F}$. (The diameter of a set U is defined as $|U| = \sup\{|x - y| : x, y \in U\}$.)
2. Suppose s is a non-negative number. For any $\delta > 0$, define

$$H_{s,\delta}(\mathcal{F}) = \inf\{\sum_{i=1}^{\infty} |U_i|^s : \{U_i\} \text{ is a } \delta\text{-cover of } \mathcal{F}\}.$$

   Then we call $H_s(\mathcal{F}) = \lim_{\delta \to 0} H_{s,\delta}(\mathcal{F})$ the outer s-dimensional Hausdorff measure of $\mathcal{F}$.
3. The Hausdorff dimension of the set $\mathcal{F}$ is defined by

$$\dim_H(\mathcal{F}) = \inf\{s : H_s(\mathcal{F}) = 0\} = \sup\{s : H_s(\mathcal{F}) = \infty\}.$$

Despite its abstract formulation, the definition relies on a rather intuitive principle. As an illustration, consider a line segment of length a. We want to cover it with s-dimensional hypercubes of side-length equal to δ. To do so most effectively we need approximately a/δ hypercubes. The Lebesgue measure of the covering is then $(a/\delta) \times \delta^s$. One can see that as $\delta \to 0$, the Lebesgue measure of the covering converges to zero if $s > 1$, and to infinity if $s < 1$. In fact, only for $s = 1$ does the measure converge to a positive real number. Therefore in this case

the Hausdorff dimension is equal to the usual Euclidian dimension of a simple line, that is, one. However, in the general case of sets with a highly irregular structure, the above calculations may yield fractional values of s.

3.2. Computing the Hausdorff dimension

Hausdorff dimensions are difficult to compute in general. There is, however, a particular class of fractals for which the task is made easier thanks to the work of Moran (1946) and Hutchinson (1981). A self-similar fractal, $\mathcal{F}$, is one in which the set itself can be partitioned, *ad infinitum*, into smaller sets, all of them having the same geometrical shape as $\mathcal{F}$. Both Koch's snowflake and the antihype fractal are self-similar. In fact, both are *attractors* of two different iterated function systems (IFS, Barnsley (1988)). The main ingredient of an IFS consists of functions that alter the size of a set without changing its shape. Such functions are called similitudes and have associated a shrinking factor or ratio r. For example, a similitude with ratio 1/3 will transform an equilateral triangle of sidelength l into a similar equilateral triangle with sidelength $l/3$. A fractal set $\mathcal{F}$ is the attractor of an IFS $(f_1, \ldots, f_n)$ with corresponding ratios $(r_1, \ldots, r_n)$ if

$$\mathcal{F} = f_1(\mathcal{F}) \cup f_2(\mathcal{F}) \cup \ldots \cup f_n(\mathcal{F}).$$

If, in addition, the IFS satisfies Moran's (1946) open set condition, that is, there exists an open set $V \subset R^d$ such that the images $\{f_j(V), 1 \leq j \leq n\}$ are disjoint subsets, then the Hausdorff dimension of $\mathcal{F}$ is the positive number s such that

$$\sum_{i=1}^{n} r_i^s = 1. \tag{3.1}$$

For Koch's snowflake, it can be shown that the Hausdorff dimension is $s = \log 4/\log 3$ (Falconer, 2003). This number is in agreement with the fact that the curve is "larger than 1-dimensional" since it has infinite length, and it is "smaller than two-dimensional" because it has zero area. In the next section we define an IFS that allows us to compute the Hausdorff dimension of our antihype fractals, a derivation that also provides an intuitive explanation of (3.1).

4. Geometric Properties of Antihype Fractals

4.1. A connection with Koch's curve

Figure 4 presents four views of the support of $\mathcal{U}_\infty^3$, that illustrate the self-similarity of $\mathcal{A}^3$. It consists of six smaller parts but with the same shape, and each of these six replicates consists of six even smaller ones, and so on. It turns out that there is a close connection between $\mathcal{A}^3$ and Koch's curve, a connection that also helps to illustrate the concept of self-similarity.

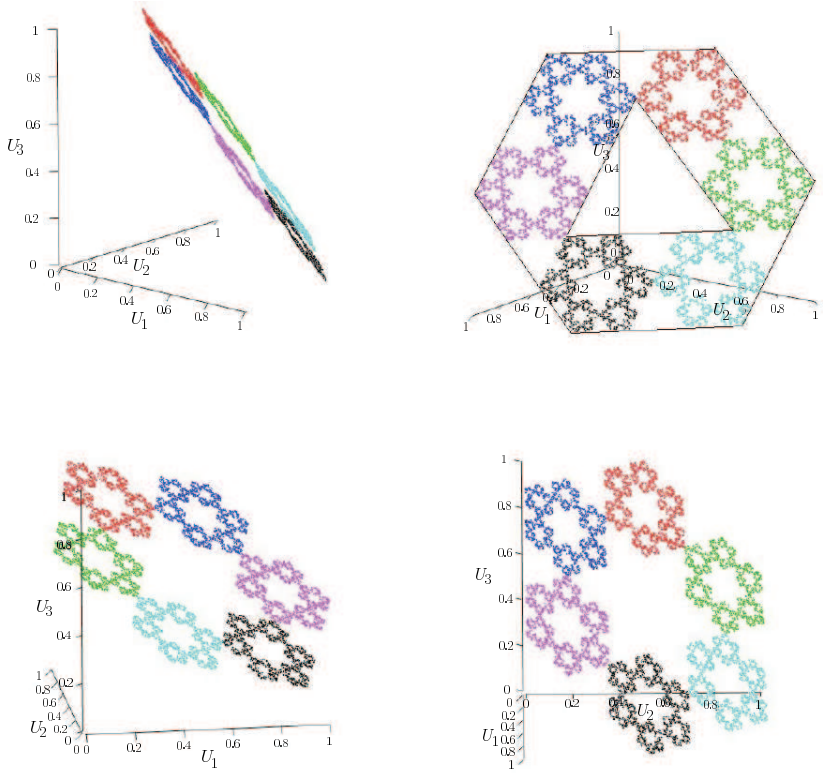

Figure 4. Four spatial views of the $\mathcal{A}^3$.

Specifically, the right image from the first row of Figure 4 shows that $\mathcal{A}^3$ inhabits a hexagon with infinitely many holes. Each of these holes apparently is bounded by a Koch's curve – seeking a rigorous proof of this assertion is too much of a detour for our tour and the task is left as a souvenir puzzle for those with inquisitive minds. Here we just use the self-similarity to prove a weaker result: the area of the largest hole is the same as the area inside the Koch's curve generated by the triangle depicted in the image. The conclusion then holds for all holes by self-similarity.

To prove this, let Δ be the area of the triangle in Figure 1 that generates the Koch's curve. By induction, at tth iteration, the snowflake curve consists of 3×4^t line segments, and the area of each triangle *to be added* to each line segment is $\Delta/9^t$. Consequently, at the t-th iteration the area inside the snowflake is $A_t = A_{t-1} + 3 \times 4^{t-1}\Delta/9^t$. This implies that the Koch's curve encloses an area equal to $A_\infty = 8\Delta/5$. Returning to Figure 4, if we denote by Δ the area of the triangle, we see that the area of the largest hole, denoted by H, is the same as $\Delta+3B$, where B is the area of any of the three regions between the three triangle's

sides and $\mathcal{A}^3$. By symmetry, B is also the area of any of the six regions between the $\mathcal{A}^3$ and the six hexagon's sides. Clearly, the area of the whole hexagon is 6Δ, and the area of the hexagon excluding the aforementioned six regions is the same as

$$H + \Big(\frac{6}{9}\Big)H + \Big(\frac{6}{9}\Big)^2 H + \cdots = 3H,$$

because every iteration of ILHS forms six smaller rosaries, each of which has the largest hole with area equal to one ninth of the area of the corresponding "parental" one. Consequently, we have $6\Delta - 6B = 3H$. Together with $H = \Delta + 3B$, this implies $H = 8\Delta/5 = A_\infty$, our desired result.

4.2. Dimension of the antihype fractals

The self-similarity of $\mathcal{A}^k$ also allows us to prove that $\mathcal{A}^k$ is invariant for a particular set of IFS that satisfies the open set condition, which in turn allows an easy calculation of its Hausdorff dimension. Specifically, for any permutation σ of $\{0, \ldots, k-1\}$, we can define a transformation, $\psi_\sigma : \mathcal{A}^k \to \mathcal{A}^k$, by

$$\psi_\sigma(x_1, \ldots, x_k) = \Big(\frac{\sigma(0) + x_1}{k}, \ldots, \frac{\sigma(k-1) + x_k}{k}\Big).$$

Each ψ_σ is a similitude with contraction ratio $1/k$, corresponding to one realization of (2.3). Indeed, $\mathcal{A}^k$ itself is just the union of all these images, as proved in Appendix:

$$\cup_\sigma \psi_\sigma(\mathcal{A}^k) = \mathcal{A}^k, \tag{4.1}$$

where the union is taken over all $k!$ possible permutations of $\{0, \ldots, k-1\}$. Evidently, all $k!$ images, $\psi_\sigma(\mathcal{A}^k)$, are identical in shape to $\mathcal{A}^k$ but with a shrinking factor $1/k$, as demonstrated in Figure 4.

Equation (4.1) implies that, if the dimension of $\mathcal{A}^k$ is s, then a Hausdorff measure H_s should satisfy

$$H_s(\mathcal{A}^k) = \sum_{i=1}^{k!} H_s(\psi_{\sigma_i}(\mathcal{A}^k)) = \sum_{i=1}^{k!} \frac{H_s(\mathcal{A}^k)}{k^s}. \tag{4.2}$$

The additivity holds here because, for $\sigma_1 \neq \sigma_2$, $\psi_{\sigma_1}(\mathcal{A}^k) \cap \psi_{\sigma_2}(\mathcal{A}^k) = \emptyset$. The last identity in (4.2) is equivalent to a change of variable formula in which the Jacobian term is from a linear transformation with contracting coefficient $1/k$ but in dimension s. This is the intuitive principle behind (3.1).

Evidently, in order for (4.2) to hold for an s such that $0 < H_s(\mathcal{A}^k) < \infty$, we must set

$$\sum_{i=1}^{k!} \frac{1}{k^s} = 1. \tag{4.3}$$

Solving (4.3) for s we deduce that the Hausdorff dimension for $\mathcal{A}^k$ is

$$\dim_H(\mathcal{A}^k) = \frac{\log(k!)}{\log(k)}. \tag{4.4}$$

4.3. Projections of antihype fractals

The above method can be used to study $\mathcal{A}^k_r$, a projection of $\mathcal{A}^k$ into one of the r-dimensional hypercubes ($2 \le r \le k$). Due to exchangeability, $\mathcal{A}^k_r$ has the same geometrical properties no matter which of the $k!/(r!(k-r)!)$ possible r-dimensional subspaces we choose to project into, as partially demonstrated in the second row of panels in Figure 4. In this case, because the dimension is reduced by only one, from $k = 3$ to $r = 2$, the number of similar pieces remains equal to six for each two-dimensional projection of $\mathcal{A}^3$. Clearly, the self-similarity of the original fractal is preserved by each of these projections. The left panel in the first row of Figure 4 shows $\mathcal{A}^3$ from a different point of view. In fact, if we moved the point of view a little to the right we would have observed only one line. This is because all the points in $\mathcal{A}^3$ are on the plane $x + y + z = 3/2$, which is just a graphical confirmation of the fact that $\mathcal{U}^3_\infty$ achieves EA.

To calculate the Hausdorff dimension of $\mathcal{A}^k_r$, we first notice that because the projection and the linear mapping ψ_σ commute, (4.1) implies that

$$\cup_\sigma \psi_\sigma(\mathcal{A}^k_r) = \mathcal{A}^k_r. \tag{4.5}$$

Unlike (4.1), however, $\psi_{\sigma_1}(\mathcal{A}^k_r) = \psi_{\sigma_2}(\mathcal{A}^k_r)$ even if $\sigma_1 \neq \sigma_2$, as long as the two permutations agree on all the components that are within the projected space. Since for each σ, there are $(k-r)!$ such equivalences, the total number of mutually exclusive $\psi_\sigma(\mathcal{A}^k_r)$ in the left-hand side of (4.5) is $k!/(k-r)!$. Consequently, (4.3) needs to be replaced by the more general

$$\sum_{i=1}^{\frac{k!}{(k-r)!}} \frac{1}{k^s} = 1, \tag{4.6}$$

which yields a generalization of (4.2) (i.e., (4.2) is special case of the following (4.7) with $r = k$),

$$\dim_H(\mathcal{A}^k_r) = \frac{\log(k!) - \log[(k-r)!]}{\log(k)}. \tag{4.7}$$

Interestingly, we note that for any *fixed* $r \ge 2$, $\dim_H(\mathcal{A}^k_r) \to r$ as $k \to \infty$. This implies that $\mathcal{A}^k_r$ becomes dense in $[0,1]^r$ while remaining a Lebesgue null set in R^r as $k \to \infty$, much like the set of all rational numbers is dense in R^1

but with Lebesgue measure zero. Figure 5 illustrates this phenomena by showing that $\mathcal{A}_2^6$ is much more dense in the unit square than $\mathcal{A}_2^3$. We also remind the reader that when $k = 2$, $\dim_H(\mathcal{A}^2) = 1$, i.e., the Hausdorff measure is the same as the Lebesgue measure on R^1. This holds intuitively because $\mathcal{U}_\infty^2$ is just the usual quantile coupling using $(U_1, 1-U_1)$, which sits on the line $x+y=1$ in R^2. This is a simple demonstration that Hausdorff dimension correctly captures the geometric properties of Lebesgue null sets.

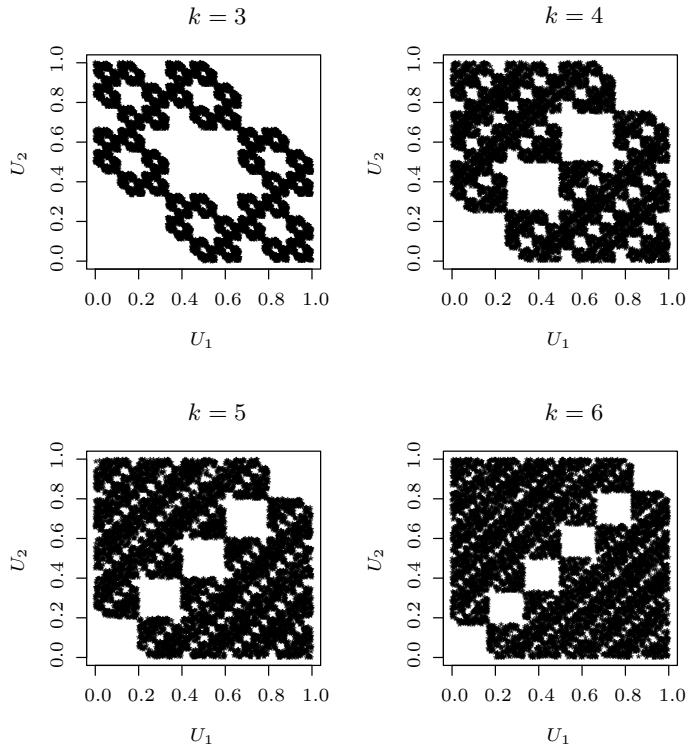

Figure 5. Coordinate two–dimensional projection of the $\mathcal{A}^k$ for different values of k.

5. A Cautionary Tale of Distributional Convergence Distances

5.1. Geometric convergence in Kolmogorov-Smirnov distance

Besides offering another connection between Monte Carlo and fractal geometry, ILHS also provides an intriguing example that is of pedagogical value. We start by considering the bivariate sequence $B_t = (U_t^{(1)}, U_t^{(2)})$, $t \geq 1$, resulting from (2.3). By the definition of (2.3), we have

$$\mathcal{U}_t^k = \frac{\mathcal{K}_t}{k} + \frac{\mathcal{K}_{t-1}}{k^2} + \cdots + \frac{\mathcal{K}_1}{k^t} + \frac{\mathcal{U}_0^k}{k^t}. \tag{5.1}$$

To properly define $\mathcal{U}_\infty^k$ and hence B_∞, we need the "time backward dual sequence"

$$\tilde{\mathcal{U}}_t^k = \frac{\mathcal{K}_1}{k} + \frac{\mathcal{K}_2}{k^2} + \cdots + \frac{\mathcal{K}_t}{k^t} + \frac{\mathcal{U}_0^k}{k^t}. \tag{5.2}$$

Clearly $\mathcal{U}_t^k = (U_t^{(1)}, \ldots, U_t^{(k)})^\top$ and $\tilde{\mathcal{U}}_t^k = (\tilde{U}_t^{(1)}, \ldots, \tilde{U}_t^{(k)})^\top$ have identical k-dimensional joint distributions for any t. The advantage of the $\tilde{\mathcal{U}}_t^k$ sequence is that $\tilde{\mathcal{U}}_t^k$ converges *point-wise*, on the infinite product space defined by $\{\mathcal{K}_t, t \geq 1\}$, to the well-defined limit $\mathcal{U}_\infty^k = \sum_{t=1}^\infty \mathcal{K}_t/k^t$. Note (2.2) is the special case of (5.2) when $k = 2$.

Now for each i, indexing the ith component of $\tilde{\mathcal{U}}_t^k$, because $\sigma_t(i-1) \leq k-1$, it is easy to show that $|U_\infty^{(i)} - \tilde{U}_t^{(i)}| \leq \sum_{j=t+1}^\infty (k-1)/k^j + 1/k^t = 2/k^t \equiv \epsilon_t$. Consequently, if we let $F_t(u,v)$ be the bivariate CDF of B_t (or $\tilde{B}_t$) for any t, including $t = \infty$, we have

$$-[F_t(u,v) - F_t(u-\epsilon_t, v-\epsilon_t)] \leq F_\infty(u,v) - F_t(u,v) \leq F_t(u+\epsilon_t, v+\epsilon_t) - F_t(u,v). \tag{5.3}$$

Because $|F(u+\alpha, v+\beta) - F(u,v)| \leq |F_x(u+\alpha) - F_x(u)| + |F_y(v+\beta) - F_y(v)|$ for any $\alpha, \beta \geq 0$, where F_x and F_y are the two marginal CDFs of F, and because marginally $U_t^{(i)} \sim$ Uniform$[0,1]$, (5.3) implies that $|F_t(u,v) - F_\infty(u,v)| \leq 2\epsilon_t = 4/k^t$. Since this bound is free of u and v, we can conclude that $\tilde{B}_t$ converges to B_∞ geometrically with respect to Kolmogorov-Smirnov (KS) distance (Kolmogorov (1933)) as $t \to \infty$, because

$$D(t,\infty) \equiv \sup_{(u,v)\in\mathbf{R}^2} |F_t(u,v) - F_\infty(u,v)| \leq \frac{4}{k^t}. \tag{5.4}$$

Incidentally, by a much more complex argument, Craiu and Meng (2005) proved a tighter bound than (5.4):

$$D(t, t+m) \leq k^{-(t-1)}(k-1)^{-(t+2)}, \quad \text{for any } t \geq 1 \text{ and } m \geq 0. \tag{5.5}$$

5.2. Non-convergence in total variation and Hellinger distances

The geometric convergence in KS distance of B_t to B_∞ is a rather strong result, which might tempt one to guess that B_t also converges under other common metrics, especially in view of the fact that (5.5) holds even when $(U_t^{(1)}, U_t^{(2)})$ is replaced by $(g_1(U_t^{(1)}), g_2(U_t^{(2)}))$ for any monotone functions $g_i, i = 1, 2$, as proved in Craiu and Meng (2005). The following general result, which should not be unexpected for an astute reader who noticed the shrinking of distributional supports, shows that such an intuition can be quite misleading.

Lemma 1. *Let $\{S_t, t = 0, 1, 2, \ldots\}$ be a sequence of μ-measurable sets in R^d such that for all t, (a) $S_{t+1} \subset S_t$, and (b) $\mu(S_{t+1}) = (1-p)\mu(S_t)$ for some t-independent constant $0 < p < 1$, where $0 < \mu(S_0) < \infty$. Let Y_t be a uniform random variable on S_t (with respect to μ). Then Y_t does not converge in the metrics defined by the Hellinger distance or L^r distances, for any $r \geq 1$.*

Proof. Without loss of generality we assume $\mu(S_0) = 1$, hence $\mu(S_t) = (1-p)^t$ and the density for Y_t is $f_t(y) = 1_{S_t}(y)(1-p)^{-t}$. The Hellinger distance between f_{t+m} and f_t is

$$H(t, t+m) \equiv \int [\sqrt{f_t(y)} - \sqrt{f_{t+m}(y)}]^2 \mu(dy) = 2\Big[1 - \int_{S_{t+m}} \sqrt{f_t(y) f_{t+m}(y)}\mu(dy)\Big]$$
$$= 2\Big[1 - \sqrt{\frac{\mu(S_{t+m})}{\mu(S_t)}}\Big] = 2[1 - \sqrt{(1-p)^m}],$$

which stays as a positive constant when $t \to \infty$. Consequently, Y_t is not a Cauchy sequence, and therefore cannot converge, in Hellinger distance.

For L^r distance, let $l_r(t, t+m) = [\int \{f_t(y) - f_{t+m}(y)\}^r \mu(dy)]^{1/r}$. Then

$$[l_r(t, t+m)]^r = \int_{S_{t+m}} \left| \frac{1}{\mu(S_t)} - \frac{1}{\mu(S_{t+m})} \right|^r \mu(dy) + \int_{S_t \cap S_{t+m}^c} \frac{1}{\mu^r(S_t)} \mu(dy)$$
$$= \frac{\mu(S_t) - \mu(S_{t+m})}{\mu^r(S_t)} \Big\{ \Big[\frac{\mu(S_t) - \mu(S_{t+m})}{\mu(S_{t+m})} \Big]^{r-1} + 1 \Big\}$$
$$= \frac{1 - (1-p)^m}{(1-p)^{(r-1)t}} \Big\{ \Big[\frac{1}{(1-p)^m} - 1 \Big]^{r-1} + 1 \Big\}.$$

This result shows that $\lim_{t\to\infty} l_r(t, t+m) = \infty$ when $r > 1$, and $\lim_{t\to\infty} l_1(t, t+m) = 2[1 - (1-p)^m] > 0$ for any $m > 0$. Hence Y_t cannot converge in terms of L^r distance for any $r \geq 1$.

If we take S_t to be the support of B_t, then conditions (a) and (b) in Lemma 1 hold with $p = k^{-1}$. Since the TV distance is half of the L_1 distance (e.g., Lindvall (1992)), Lemma 1 implies that B_t does not converge in TV distance, Hellinger distance or any L^r distance with $r \geq 1$, even though it converges geometrically in the Kolmogorov-Smirnov distance. (Obviously, this lemma is applicable to $\mathcal{U}_t^k$ itself, as well as to any uniform random variables on supports formed by the iterative process of generating any self-similar fractal contained in a μ-finite set, where μ does not have to be Lebesgue measure.) However, for many practical applications in Monte Carlo, the distribution of B_t is very close to the distribution of B_∞, even for $t \approx 5$ (see Craiu and Meng (2005)). This raises the question of whether convergence in the TV distance is a too stringent criteria for measuring convergence in some practical situations. For

example, had we insisted on using the TV distance, we might have abandoned ILHS because we would have wrongly believed that EA can never be effectively achieved with any finite number of iterations. This question is worth further investigation because TV distance is the standard metric for many theoretical results regarding convergence in MCMC (e.g., Rosenthal (1995); Marchev and Hobert (2004)). See Gibbs and Su (2001) for more comparisons between different metrics and possible applications to MCMC.

6. An End-of-tour Souvenir

In practice, the $\mathcal{U}_t^k$ generated by ILHS will be part of the arguments for some general functions. In general, the negative correlations may not be preserved even if these functions are monotone. However, the $\mathcal{U}_t^k$ generated by ILHS has the nice property called *negative association* (Joag-Dev and Proschan (1983)), which means that the non-positiveness property of the correlation is preserved under monotone transformations; see Craiu and Meng (2005) for details. In order to extend this property to $\mathcal{U}_\infty^k$, we need to show that any component-wise monotone (all in the same direction) function on R^k is almost surely continuous with respect to a (non-trivial) Hausdorff measure on the support of $\mathcal{U}_\infty^k$. Although it is quite well-known that monotone functions on R^k are almost surely continuous with respect to Lebesgue measure, whether this result holds for a Hausdorff measure on $\mathcal{A}^k$ is currently unclear. We would certainly invite anyone who is interested in such theoretical investigations to take this puzzle home after leaving our tour. If the result holds, then ILHS indeed has all the desirable properties anticipated in Craiu and Meng (2005). If not, then there would be yet another "tourist attraction" for this tour, one that will show how easily our intuition can mislead us in the strangely fascinating world of fractals.

Acknowledgement

We would like to thank Steven Lalley and Michael Wichura for helpful exchanges and three reviewers for constructive comments. This research was supported in part by several NSF, NSA, and NSERC grants.

Appendix

We prove that if P_k is the set of all possible permutations of $\{0, 1, \ldots, k-1\}$, then

$$\cup_{\sigma \in P_k} \psi_\sigma(\mathcal{A}^k) = \mathcal{A}^k. \tag{A.1}$$

First, by the construction of ILHS, we have the representation

$$\mathcal{A}^k = \Big\{ \tilde{x} = (x_1, \ldots, x_k)^\top : x_j = \sum_{t=1}^{\infty} \frac{\sigma_t(j-1)}{k^t} \Big\}, \tag{A.2}$$

where σ_t are i.i.d. permutations of $\{0, 1, \ldots, k-1\}$.

To prove (A.1), we first note that for any $\tilde{x} \in \mathcal{A}^k$, we can let $\tilde{y} = (y_1, \ldots, y_k)$ with

$$y_j = \sum_{t=1}^{\infty} \frac{\sigma_{t+1}(j-1)}{k^t}.$$

Clearly, by (A.2), $\tilde{y} \in \mathcal{A}^k$ because $(\sigma_{t+1}(0), \ldots, \sigma_{t+1}(k-1))$, $t = 1, 2, \ldots$ are also i.i.d. permutations of $\{0, \ldots, k-1\}$. Furthermore, if we let $\sigma^* = (\sigma_1(0), \ldots, \sigma_1(k-1))$ then $\psi_{\sigma^*}(\tilde{y}) = \tilde{x}$ by our construction. Therefore $\mathcal{A}^k \subset \cup_{\sigma \in P_k} \psi_\sigma(\mathcal{A}^k)$.

On the other hand, if $\tilde{x} \in \cup_{\sigma \in P_k} \psi_\sigma(\mathcal{A}^k)$, then there exists $\tilde{y} \in \mathcal{A}^k$ such that for at least one $\sigma \in P_k$, we have $\tilde{x} = \psi_\sigma(\tilde{y})$. Using the characterization of the set $\mathcal{A}^k$ given in (A.2), $\tilde{x} = \psi_\sigma(\tilde{y}) \in \mathcal{A}^k$.

References

Barnsley, M. F. (1988). *Fractals Everywhere*. Academic, Boston.

Chatterjee, S. and Yilmaz, M. R. (1992). Chaos, fractals and statistics. *Statist. Sci.* **7**, 49-121.

Craiu, R. V. and Meng, X. L. (2005). Multi-process parallel antithetic coupling for backward and forward MCMC. *Ann. Statist.* **33**, 661-697.

Edgar, G. A. (1990). *Measure, topology, and fractal geometry*. Springer-Verlag.

Falconer, K. (2003). *Fractal Geometry – Mathematical Foundations and Applications*. Second edition. Wiley, New York.

Gerow, K. and Holbrook, J. (1996). Statistical sampling and fractal distributions. *Math. Intelligencer* **18**, 12-22.

Gibbs, A. L. and Su, F. E. (2001). On choosing bounding probability metrics. *Internat. Statist. Rev.* **70**, 419-435.

Hausdorff, F. (1919). Dimension und äusseres mass. *Math. Ann.* **79**, 157-179.

Hutchinson, J. E. (1981). Fractals and self-similarity. *Indiana Univ. Math. J.* **30**, 713-747.

Iman, R. L. (1999) Latin hypercube sampling. In *Encyclopedia of Statistical Sciences*, Wiley, New York.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Ann. Statist.* **11**, 286-295.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Instituto Italiano degli Attuari*, **4**, 83-91.

Kostelich, E. J. and Yorke, J. A. (1988). Practical considerations in estimating dimension from time series data. *Phys. Rev. A* **38**, 1649-1652.

Lindvall, T. (1992). *Lectures on the Coupling Method*. Wiley, New York.

Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. Freeman and Company, New York.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *J. Amer. Statist. Assoc.* **99**, 228-239.

McKay, M. D. and Beckman, R. J. and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239-245.

Moran, P. A. P. (1946). Additive functions of intervals and Hausdorff measure. *Math. Proc. Cambridge Philos. Soc.* **42**, 15-23.

Moran, P. A. P. (1967). Testing for correlation between non-negative variates. *Biometrika* **54**, 385-394.

Owen, A. B. (1992). A central limit theorem for Latin hypercube sampling. *J. Roy. Statist. Soc. Ser. B* **54**, 541-551.

Palmore, J. (1995). Chaos, entropy and integrals for discrete dynamical systems on lattices. *Chaos Solitons Fractals* **8**, 1397-1418.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90**, 558-566.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29**, 143-151.

Department of Statistics, University of Toronto, 100 St. George St., Toronto, Ontario, M5S 3G3, Canada.

E-mail: craiu@utstat.toronto.edu

Department of Statistics, Harvard University, Cambridge, MA 02138-2901, U.S.A.

E-mail: meng@stat.harvard.edu

(Received June 2004; accepted January 2005)