Zeitschrift: Elemente der Mathematik
Herausgeber: Schweizerische Mathematische Gesellschaft
Band: 51 (1996)

Artikel: Chance and chaos
Autor: Barbour, A.D.
DOI: https://doi.org/10.5169/seals-46958

El. Math. *51* (1996)

0013-6018/96/020057-07$1.50 + 0.20/0

Chance and Chaos

A. D. Barbour

Andrew Barbour wurde 1947 geboren. Er studierte Mathematik und Statistik an der University of Cambridge, wo er 1974 promovierte und ein Fellowship am Gonville and Caius College antrat. Seit 1983 ist er Professor für Biomathematik an der Universität Zürich. Er interessiert sich in erster Linie für angewandte Stochastik; daneben hat er auch über Mathematik der Epidemiologie, insbesondere über die mathematische Modellierung der Verbreitung von parasitischen Krankheiten gearbeitet.

In recent years, the idea of chaos produced by simple deterministic algorithms has come to rival the classical concept of randomness as a model for the disorder observed in nature. Indeed, the difficulty in distinguishing between them is happily exploited for the generation of random numbers in computers. In this article, we show that the two can at times be seen as different aspects of the same phenomenon, in which case randomness can be used to explain chaos.

The example we consider is that of the map $h:\ [0,1] \to [0,1]$ defined by

$$h(x) = \min\{x/c, (1-x)/(1-c)\},$$

Es gibt heutzutage zwei verschiedene Modelle, mit denen wir versuchen, das Unvorhergesehene zu beschreiben. Das erste, das 1494, also vor genau 500 Jahren in der Arbeit von Pacioli entstand, jedoch erst 1933 von Kolmogorov die moderne Axiomatisierung bekam, besteht aus dem üblichen Begriff des "Zufalls". Das zweite Modell beruht auf der Erkenntnis, dass die Lösungen von vielen Gleichungssystemen, die in der Physik auftauchen, äusserst empfindlich auf kleine Änderungen in den Anfangswerten reagieren, so dass genaue Prognosen über längere Zeit unrealistisch präzise Information zum Anfangszustand benötigen würden. Im Prinzip sind solche Phänomene genau vorhersehbar, in der Praxis jedoch nur für eine beschränkte Zeit; die Wetterprognose bietet ein wohlbekanntes Beispiel dafür. Obwohl im zweiten Modell der Zufall im ursprünglichen Sinne gar nicht erscheint, haben die Abläufe in einem solchen "deterministischen Chaos" viele Eigenschaften gemeinsam mit denjenigen von echten Zufallsexperimenten. In diesem Beitrag wird ein System untersucht, wo sich die beiden Auffassungen gleichzeitig als gültig erweisen. *adb*

for some $c \in (0, 1)$. Starting with any $x_0 \in [0, 1]$, we recursively define $x_n = h(x_{n-1})$, writing $x_n = h_n(x_0)$ for short. Then it turns out that the set of values $\{x_0, x_1, \ldots, x_n\}$, ignoring the order of their generation, almost always looks statistically very like a random sample from the uniform distribution on $(0, 1)$.

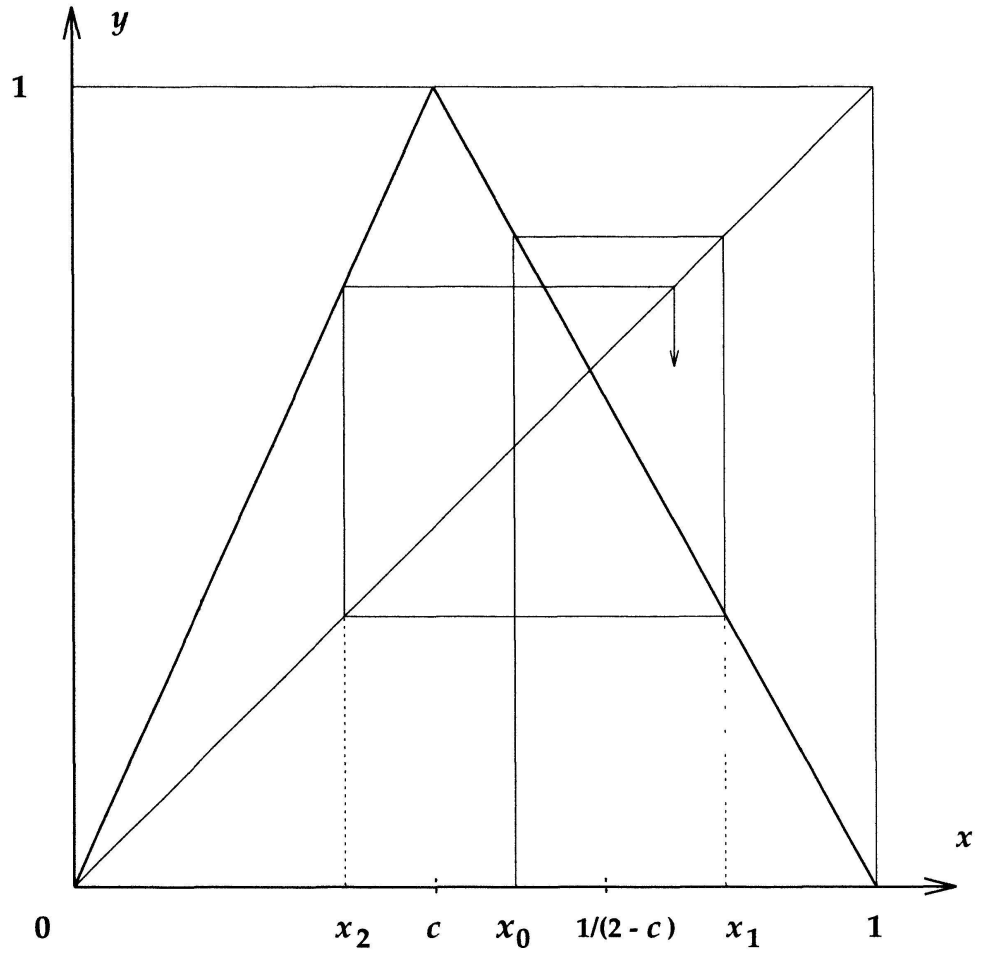

Fig. 1 The function $h(x)$ and the beginning of the sequence $x_0, x_1, x_2, \ldots$.

This is at first sight surprising. One might have expected that x_n would converge to one of the two fixed points 0 and $1/(2 - c)$ of h, or perhaps assume some periodic behaviour, but it does not: for any x_0 outside a small exceptional set E, which includes the fixed points of h and all their pre-images (for example, $h_2(c) = h(1) = 0$), all the many periodic points of h, and so on, the sequence x_n keeps on hopping around, leaving apparently random footprints on $(0, 1)$. Try it on your computer, but don't take $c = 1/2$! And the exceptional set E is so small that, if x_0 were chosen at random from the uniform distribution on $(0, 1)$, the chance that $x_0 \in E$ is zero.

In order to account for this apparent randomness, we shall explore in more detail one of the key ingredients in one's intuitive idea of chaos, that of unpredictability. A sequence $x_0, x_1, \ldots$ might be said to be (asymptotically) unpredictable if knowledge of x_0 gives (little or) no information about the value of x_n (for large n). Clearly, the sequence $x_n = h_n(x_0)$ is not unpredictable in this sense, since, given x_0, each x_n can in principle

be computed exactly. However, this argument presupposes that x_0 is known precisely, whereas, in most practical circumstances, this will not be the case: limitations of measurement reduce one's knowledge at best to being certain that x_0 is in a known interval, say $(d-\delta, d+\delta)$, for some small δ. What we shall now show is that, for large n, this information is actually of almost no help at all in predicting x_n.

To express this fact cleanly, let $1_{\{y\in A\}}$ be the function taking the value 1 if $y \in A$ and 0 otherwise. Then

$$\left|\frac{1}{2\delta}\int_{d-\delta}^{d+\delta} 1_{\{a<h_n(x)\le b\}}\,dx - (b-a)\right| \le 2\alpha^n\delta^{-1}(b-a), \tag{1}$$

for any values $0 \le a < b \le 1$, where $\alpha = \max\{c, 1-c\} < 1$. In other words, the fraction of starting points x_0 in the interval $(d-\delta, d+\delta)$ for which $a < h_n(x_0) \le b$ is for large n very close to $(b-a)$. But since $(b-a)$ is just the probability that a randomly chosen point of $(0,1)$ would be in the interval (a,b), the estimate (1) confirms that, for large n, the starting information that $d-\delta < x_0 < d+\delta$ is of almost no help in predicting the value of x_n: irrespective of the value of d, a blind guess at the value of x_n does (almost) just as well as using the information that $x_n = h_n(x_0)$. Thus the recursion $x_n = h(x_{n-1})$, in this specific sense, generates "unpredictable" sequences. What is more, the predictive value of the initial information is seen to be at most of order $\delta^{-1}\alpha^n$, decaying geometrically fast with n, so that forecasting further ahead by increasing the precision of the starting data is ineffective: doubling the precision by halving δ achieves a forecast with the previous accuracy only for an extra $\log 2/\log(1/\alpha)$ steps.

Note, in passing, that rounding error in the computer introduces a little unexpected "uncertainty" into any attempt to compute $h_n(x_0)$ numerically, so that the answer obtained may be surprising: try $c = 1/2$!

How should we prove the unpredictability (1)? Here, we shall do so using the properties of some simple random sequences, known as Markov chains. A random sequence $X_0, X_1, \ldots$ is called a *Markov chain* if, given the knowledge that $X_0 = x_0$, $X_1 = x_1, \ldots, X_{j-1} = x_{j-1}$ and $X_j = x$, the distribution of X_{j+1} depends only on x, and not on j or on the "past history" $x_0, \ldots, x_{j-1}$ of the chain. Our sequence $\{x_n,\ n \ge 0\}$ is a Markov chain, since, with all this information, we know that $X_{j+1} = h(x)$ is determined by x alone – the example is degenerate, in the sense that the distribution of X_{j+1} assigns probability 1 to the single possible outcome $h(x)$, so that there is no real randomness at all, but never mind: it is still a (purely deterministic) Markov chain. Now a Markov chain is called *stationary* if the distribtuion of the (random) quantity X_j is the same for all j. This is clearly not normally the case for our sequence $\{x_n\}$ if $X_0 = x_0$ is fixed, because then $X_1 = h(x_0) \ne x_0$ (if $x_0 \notin \{0, 1/(2-c)\}$), and so X_1 does not have the same distribution as X_0: each takes a single certain value, and they are different. However, if X_0 is chosen uniformly at random from $(0,1)$ and $X_n = h_n(X_0)$, the sequence $\{X_n\}$ *is* stationary. This is because

$$h^{-1}(a,b) = (ca, cb) \cup \big(1-(1-c)b, 1-(1-c)a\big),$$

and hence, if Z is uniformly distributed on $(0,1)$,

$$\begin{aligned}\mathbb{P}[a < h(Z) \leq b] &= \mathbb{P}[Z \in h^{-1}(a,b)] \\ &= (cb - ca) + (1 - (1-c)a) - (1 - (1-c)b) \\ &= c(b-a) + (1-c)(b-a) = b - a \\ &= \mathbb{P}[a < Z \leq b],\end{aligned} \tag{2}$$

so that $h(Z)$ is also uniformly distributed on $(0,1)$ – now use induction!

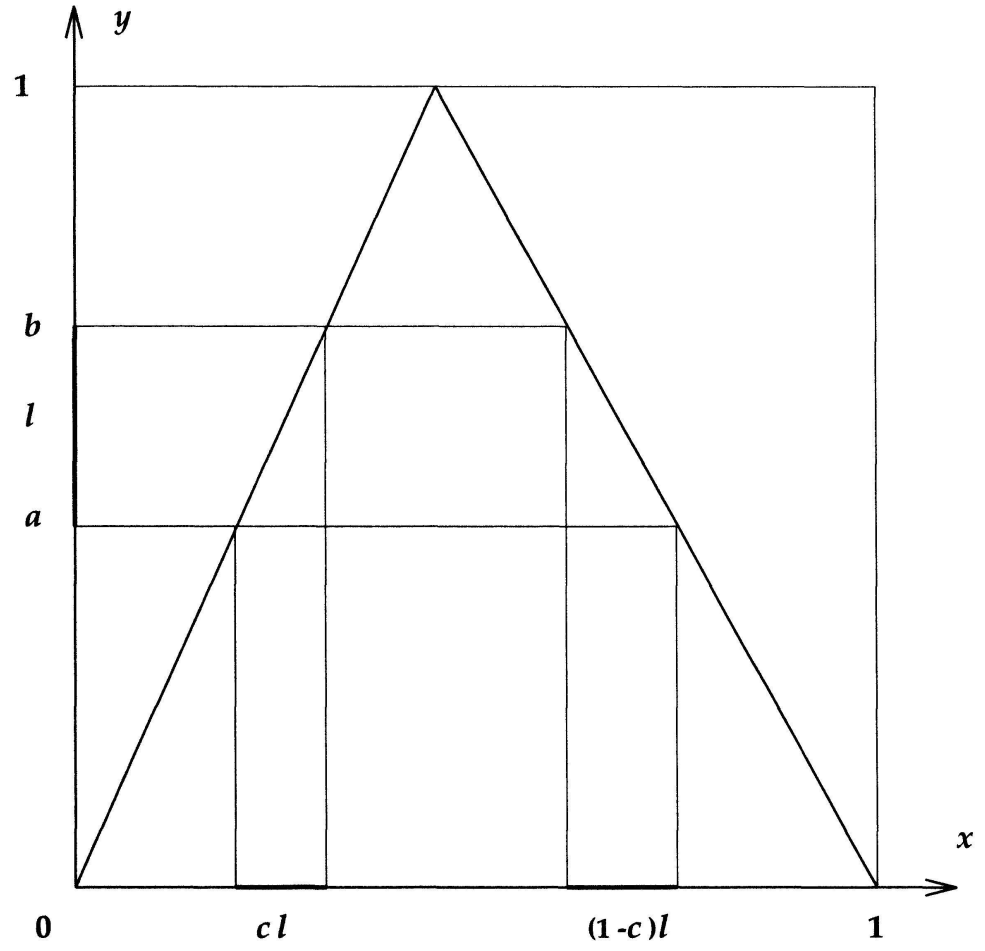

Fig. 2 The image of (a,b) under h^{-1}.

A key feature of stationary Markov chains is that, if they are viewed backwards in time, they are again stationary Markov chains, though usually with a different transition matrix. That is, if we know that $X_j = x$ and $X_{j+1} = x_{j+1}, \ldots, X_n = x_n$, and want to speculate on the most recent unknown value X_{j-1}, then the conditional distribution of X_{j-1} given all this information once again depends only on x, and not on j or on $x_{j+1}, \ldots, x_n$. Perhaps surprising, but the calculation with conditional probabilities required to see that this is true is actually very simple. So what happens if we time-reverse the stationary chain $X_n = h_n(X_0)$? The segment $(Y_0, Y_1, \ldots, Y_n)$ of the time reversed chain has to have the same distribution as a segment of the X-chain in reverse order, say $(X_n, X_{n-1}, \ldots, X_0)$, so that immediately each Y_j has to have the uniform distribution on $(0,1)$. All that is

then needed is the joint distribution of a consecutive pair, say (Y_0, Y_1), to determine the distribution of Y_1 for each value y of Y_0, and hence, by induction, to specify the whole Y-chain. Now the distribution of the pair (Y_0, Y_1) is the same as that of the pair (X_1, X_0) (time reversal), which is concentrated on the set $H = \{(h(x), x),\ x \in (0,1)\}$. Thus, given any $y \in (0,1)$, there are two points (y, cy) and $(y, 1-(1-c)y)$ in H which have first coordinate y, so that, given $Y_0 = y$, there are just *two* possible choices (y and $1-(1-c)y$) for Y_1, and the Y-chain will have been specified when the probabilities associated with the two choices have been found. We claim that

$$\mathbb{P}[Y_1 = cy \mid Y_0 = y] = c; \qquad \mathbb{P}[Y_1 = 1-(1-c)y \mid Y_0 = y] = 1-c, \tag{3}$$

for all y. To check this, observe that (X_1, X_0) gives probability $b-a$ to the set $H_{a,b} = \{(h(x), x),\ a < x \le b\} \subset H$, because X_0 is uniform on $(0,1)$. If $0 \le a < b < c$, we can rewrite $H_{a,b}$ as $\{(y, cy),\ a/c < y \le b/c\}$. Because Y_0 is uniform on $(0,1)$, Y_0 gives probability $(b/c) - (a/c) = (b-a)/c$ to $\{y :\ a/c < y \le b/c\}$, and if (3) is true it follows that (Y_0, Y_1) gives probability $((b-a)/c) \cdot c = (b-a)$ to $H_{a,b}$, as required by time reversal. A similar argument works for $c < a < b \le 1$. Thus we have the following

Reversal Theorem. *The reverse ordered segment $(X_n, X_{n-1}, \ldots, X_0)$ of the stationary X-chain has the same distribution as $(Y_0, Y_1, \ldots, Y_n)$, where Y_0 is uniformly distributed on $(0,1)$ and, conditional on $Y_j = y$, Y_{j+1} takes the value cy with probability c and the value $1-(1-c)y$ with probability $1-c$.*

The interest of the Reversal Theorem is that features of a chain Y in which randomness keeps occurring at every step can be used to deduce results about the essentially deterministic X-sequence: a "chaotic" sequence is mirrored by a "random" sequence.

We now make a construction known as a *coupling* of Y-chains. Take any two starting points $Y_0^{(1)}$ and $Y_0^{(2)}$ in $(0,1)$, and define sequences $(Y_0^{(1)}, \ldots, Y_n^{(1)})$ and $(Y_0^{(2)}, \ldots, Y_n^{(2)})$ recursively as follows. Given the values $(Y_0^{(1)}, \ldots, Y_j^{(1)})$ and $(Y_0^{(2)}, \ldots, Y_j^{(2)})$, choose at random L, with probability c, or R, with probability $1-c$. If L, define

$$Y_{j+1}^{(1)} = cY_j^{(1)} \quad \text{and} \quad Y_{j+1}^{(2)} = cY_j^{(2)};$$

if R, define

$$Y_{j+1}^{(1)} = 1-(1-c)Y_j^{(1)} \quad \text{and} \quad Y_{j+1}^{(2)} = 1-(1-c)Y_j^{(2)};$$

so that, in either case,

$$|Y_{j+1}^{(1)} - Y_{j+1}^{(2)}| \le \alpha |Y_j^{(1)} - Y_j^{(2)}|,$$

with $\alpha = \max\{c, 1-c\}$ as above. This construction yields our next result.

Coupling Theorem. *If $Y_0^{(1)}$ and $Y_0^{(2)}$ are chosen independently at random from the uniform distribution on $(0,1)$, then each of the sequences $(Y_0^{(1)}, \ldots, Y_n^{(1)})$ and $(Y_0^{(2)}, \ldots, Y_n^{(2)})$ is a stationary Y-chain, and hence by the Reversal Theorem has the same distribution as that of $(X_n, X_{n-1}, \ldots, X_0)$ in the stationary X-chain. Furthermore, by construction,*

$$|Y_n^{(1)} - Y_n^{(2)}| \le \alpha^n |Y_0^{(1)} - Y_0^{(2)}| \le \alpha^n,$$

and $Y_0^{(1)}$ is independent of $Y_n^{(2)}$.

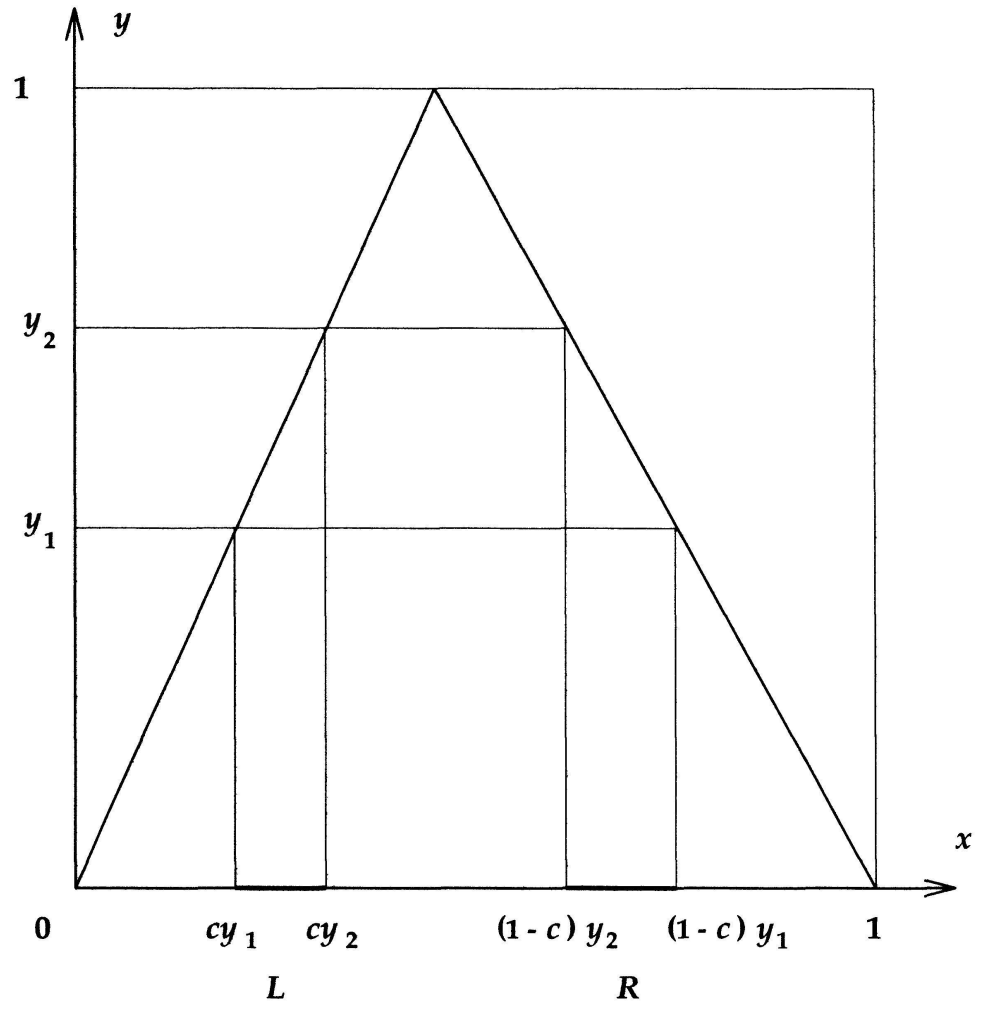

Fig. 3 The coupling construction.

What does this arcane construction tell us? The Reversal Theorem tells us that the sequence $(X_n, X_{n-1}, \ldots, X_0)$ behaves just like the sequence $(Y_0, Y_1, \ldots, Y_n)$. The coupling construction tells us that the *randomness* in the step by step choices of the Markov chain Y determines the value of Y_n to within an error of $\pm\alpha^n$, *irrespective* of the value of Y_0. Thus the distribution of Y_n is the same, modulo errors of order α^n, whatever the value of Y_0, and hence, in this sense, Y_n and Y_0 are effectively independent. By time reversal, we thus have X_0 and X_n effectively independent, implying the unpredictability of X_n when X_0 is (nearly) known.

That this is more than just a heuristic is shown by the completion of the proof of (1). Observe that, in probabilistic notation,

$$I_{a,b} = \frac{1}{2\delta}\int_{d-\delta}^{d+\delta} 1_{\{a<h_n(x)\le b\}}\,dx = \frac{1}{2\delta}\mathbb{E}\left[1_{\{X_0\in(d-\delta,d+\delta)\}}1_{\{X_n\in(a,b]\}}\right]$$

is just an expectation involving X_0 and X_n. Therefore, by the Reversal Theorem,

$$I_{a,b} = \frac{1}{2\delta}\mathbb{E}\left[1_{\{Y_0^{(1)}\in(a,b]\}}1_{\{Y_n^{(1)}\in(d-\delta,d+\delta)\}}\right],$$

also. Now, for any y_1, y_2 and (u, v), the difference $|1_{\{y_1 \in (u,v)\}} - 1_{\{y_2 \in (u,v)\}}|$ is either 0 or 1, and can only take the value 1 if y_1 and y_2 are on opposite sides of one of the end points u or v. Thus, if $|y_1 - y_2| < \varepsilon$,

$$|1_{\{y_1 \in (u,v)\}} - 1_{\{y_2 \in (u,v)\}}| \leq 1_{\{y_2 \in (u-\varepsilon, u+\varepsilon) \cup (v-\varepsilon, v+\varepsilon)\}}.$$

Putting $u = d - \delta$, $v = d + \delta$ and $\varepsilon = \alpha^n$, it follows that

$$|1_{\{y_1 \in (d-\delta, d+\delta)\}} - 1_{\{y_2 \in (d-\delta, d+\delta)\}}| \leq 1_{\{y_2 \in G\}}, \tag{4}$$

with

$$G = (d - \delta - \alpha^n, d - \delta + \alpha^n) \cup (d + \delta - \alpha^n, d + \delta + \alpha^n).$$

Hence, using the Coupling Theorem to replace $Y_n^{(1)}$ by $Y_n^{(2)}$ in the formula for $I_{a,b}$, with an error bounded because of (4), we find that

$$\begin{aligned} \left| I_{a,b} - \frac{1}{2\delta} \mathbb{E}\left[1_{\{Y_0^{(1)} \in (a,b]\}} 1_{\{Y_n^{(2)} \in (d-\delta, d+\delta)\}} \right] \right| \\ \leq \frac{1}{2\delta} \mathbb{E}\left[1_{\{Y_0^{(1)} \in (a,b]\}} 1_{\{Y_n^{(2)} \in G\}} \right] \end{aligned} \tag{5}$$

But now, by the Coupling Theorem, $Y_0^{(1)}$ and $Y_n^{(2)}$ are *independent* and uniform on $(0, 1)$, so that we can compute all the expectations in (5) exactly, giving

$$|I_{a,b} - \frac{1}{2\delta}(b - a).2\delta| \leq \frac{1}{2\delta}(b - a).4\alpha^n,$$

as claimed in (1).

The Reversal and Coupling Theorems can be used for much more than proving (1); almost any "random" feature of the sequences $x_0, x_1, \ldots, x_n$ can be derived from them. Nor is the method restricted only to functions h of the form we have considered here, though, for more complicated functions, the technicalities can obscure the essential simplicity of the argument. But the real aim of the paper is to illustrate three more fundamental facts; that, in this example, the difference between chaos and randomness is merely a difference of viewpoint; that methods from one branch of mathematics can find fruitful application in another; and that the "coupling method" in the theory of stochastic processes is a wonderful tool for constructing simple and illuminating proofs.

A. D. Barbour
Abteilung für angewandte Mathematik
Universität Zürich
Winterthurerstr. 190
CH-8057 Zürich