

Abstraction vs. Application: Itô's Calculus, Wiener's Chaos, and Poincaré's Tangle

Daniel L. Goroff

November 26, 2015

Not necessarily views of the Sloan Foundation.

Professor Kiyosi Itô

- A culminating hero in the story of how probability became part of mathematics.
- A tale that includes: scandal (Poincaré); eccentricity (Wiener); and music (Itô).



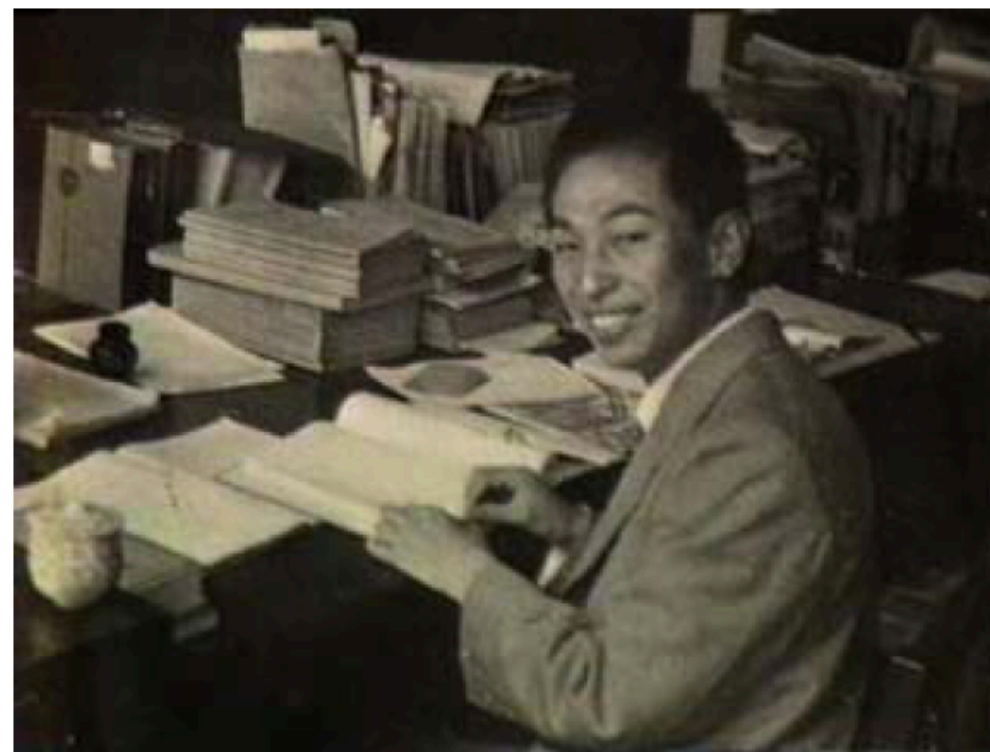
Part of Mathematics?

- Poincaré 1908: “Mathematics is the art of calling different things by the same name.”
- Poincaré 1912: “One can scarcely give a satisfactory definition of probability.”
- von Mises 1919: “In fact, one can scarcely characterize the present state other than that probability is not a mathematical discipline.”
- Even Itô wrote that, “[At university], I doubted whether probability was an authentic mathematical field.”
- By the end of his career, no one doubted that Professor Itô was a world-class and celebrated mathematician whose field was probability!

Probability is Hard

- Hilbert's Sixth Problem had specifically called for an axiomatization of Probability (and Mechanics).
- Poincaré 1897: When learning mathematics, ontogeny recapitulates phylogeny.
- Historical development of geometry vs. probability
- Built on two traditional but problematic principles: equal probabilities and small probabilities (Cournot).
- Basics: Whatever is a random variable?
- Russell 1918: "Mathematics can be defined as the subject where we never know what we are talking about nor whether what we are saying is true."

Itô wrote: “Soon after joining the Statistics Bureau of the Cabinet Secretariat, when I was still grappling with the question of how to define the random variable in probability theory, I found a book written by the Russian mathematician Kolmogoroff. Realizing that this was exactly what I had been looking for, I read through the book in one sitting. In *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Basic Concepts of Probability Theory), written in German in 1933, Kolmogoroff attempted to define random variables as functions in a probability space, and to systematize the theory of probability in terms of the theory of measures. I felt as if this book cleared the mist that was blocking my vision, leading me to finally believe that probability theory can be established as a field of modern mathematics.”

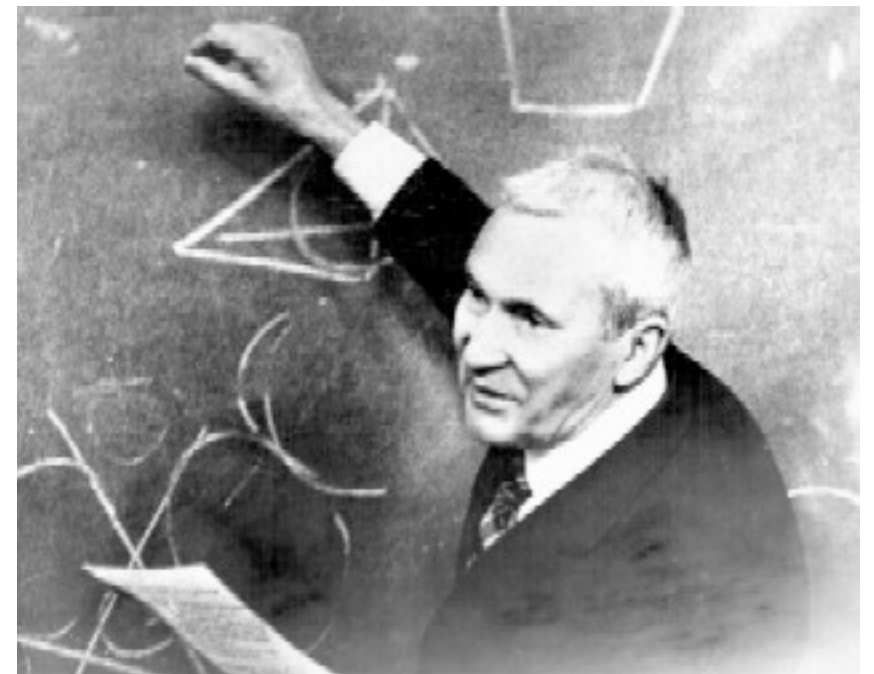


Kolmogoroff's 1933 Probability Axioms

- Measurable Space is a pair (S, Σ) , where S is a set and Σ is a sigma-field, i.e., a collection of subsets that includes \emptyset and S , and that is closed under countable set operations.
- Probability Space is a triple (Ω, \mathcal{F}, P) where (Ω, \mathcal{F}) is a measurable space and P is a probability measure, i.e., a nonnegative and countably additive set function on the measurable space with total mass one.
- Elements $A \in \mathcal{F}$ are called measurable sets and thought of as events. The measure assigns to each a number $P(A)$ between zero and one that we interpret as the probability of that event. Note, e.g., that $A \subset B \Rightarrow P(A) \leq P(B)$.

Kolmogoroff's Take on Probability

- A random variable is a measurable function X from a probability space (Ω, \mathcal{F}, P) to a measure space (S, Σ) called the state space. This just means that the inverse image of a measurable set in the state space is an event in the probability space and so can be assigned a measure.
- Defined expectation as an integral $E(X) = \int_{\Omega} X(\omega) dP(\omega)$
- Defined conditional probability as a derivative $P(A|B)$
- Nature of Ω left mysterious.
States of the world?
Place to draw Venn diagrams?



Doob wrote:

“It was a shock for probabilists to realize that a function is glorified into a random variable as soon as its domain is assigned a probability distribution with respect to which the function is measurable.

In a 1934 class discussion of bivariate normal distributions Hotelling remarked that zero correlation of two jointly normally distributed random variables implied independence, but it was not known whether the random variables of an uncorrelated pair were necessarily independent. Of course he understood me at once when I remarked after class that the interval $[0, 2\pi]$ when endowed with Lebesgue measure divided by 2π is a probability measure space, and that on this space the sine and cosine functions are uncorrelated but not independent random variables. He had not digested the idea that a trigonometric function is a random variable relative to any Borel probability measure on its domain.

The fact that nonprobabilists commonly denote functions by f, g , and so on whereas probabilists tend to call functions random variables and use the notation X, Y and so on at the other end of the alphabet helped to make nonprobabilists suspect that mathematical probability was hocus pocus rather than mathematics. And the fact that probabilists called some integrals ‘expectations’ and used the letters E or M instead of integral signs strengthened the suspicion.”

Stochastic Processes

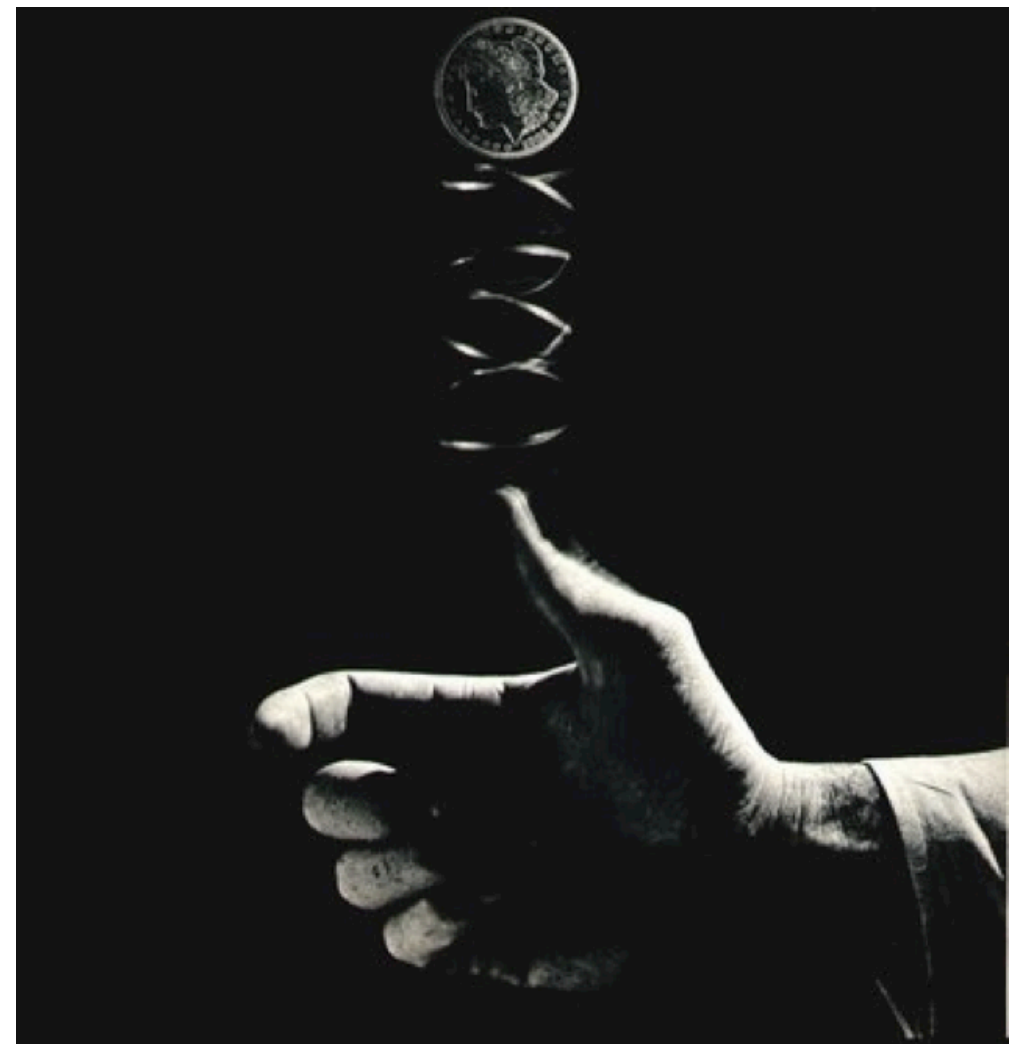
- Axioms are a rhetorical contribution, not research. Itô especially needed these axioms to define and work on the theory of stochastic processes.
- As we now understand, each is a just a collection of random variables indexed by a set T (usually time). For example, think of successive coin flips $\{X_1, X_2, X_2, \dots\}$.
- Sigma-field is generated by finite subsets of those random variables to make them measurable. Can define measure on these “cylinder sets.” Kolmogoroff showed this extends to a measure on the whole sigma-field generated.
- Probability Space of all sequences of H’s and T’s with a shift map as the passage of time called a Bernoulli Process.

Coin Flip Model

- So the n^{th} flip is a measurable function

$$X_n : (\Omega, \mathcal{F}, P) \rightarrow (S = \{H, T\}, \Sigma)$$

- Think of Tyche drawing $\omega \in \Omega$ that determines $X_n(\omega)$.
- Goethe: “Mathematicians are like Frenchmen: whatever you say they translate into their own language and henceforth it means something entirely different.”
- Not so natural or intuitive?



Tversky and Kahneman

Linda is 31 years old, single, outspoken, and bright. At college, she majored in philosophy and was concerned with discrimination, social justice, and anti-nuclear rallies.

Rank these possibilities from one (most likely) to five (least):

___ Linda is a teacher.

___ Linda works in a bookstore and takes yoga.

___ Linda is a bank teller.

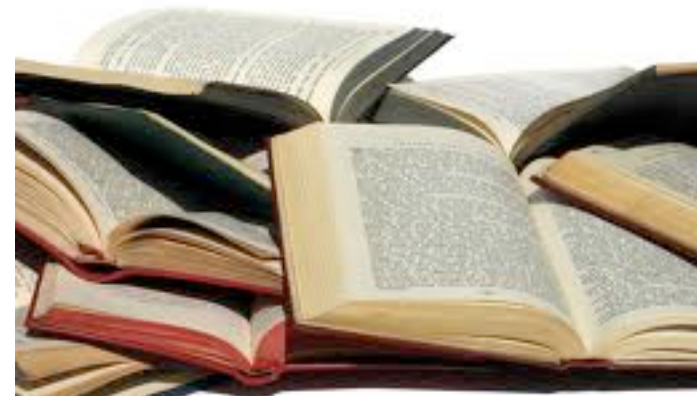
___ Linda sells insurance.

___ Linda is a bank teller and is active in the feminist movement.

Word Problems

- In the first five pages of a typical English language novel, how many words six letter words would you expect to find with the penultimate letter n? I.e., of the form:

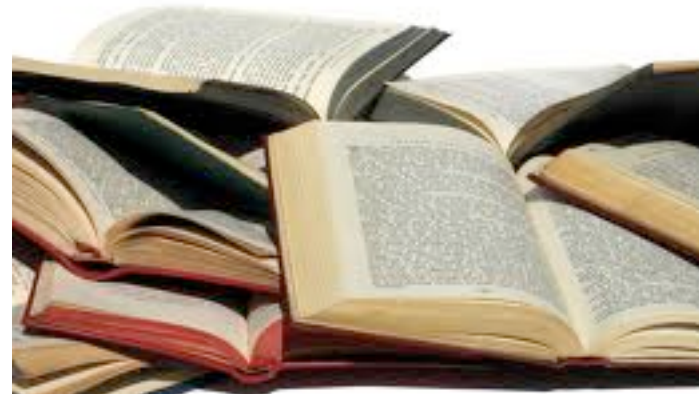
_ _ _ _ n _



Word Problems

- In the first five pages of a typical English language novel, how many six letter words would you expect to find with the penultimate letter n? I.e., of the form:

— — — — n —



- In the first five pages of a typical English language novel, how many words six letter words would you expect to find whose last letters are ing? I.e., of the form:

— — — i n g

Salesman Problem

- Tom is either a Salesman or a Librarian.
- His personality has been described as Quiet.
- Which is more likely, S or L?



Salesman Problems

- Tom is either a Salesman or a Librarian.
- His personality has been described as Quiet.
- Which is more likely, S or L?



- Fred is either a Salesman or Librarian?

Salesman Problems

- Tom is either a Salesman or a Librarian.
- His personality has been described as Quiet.
- Which is more likely, S or L?



- Fred is either a Salesman or Librarian?
- $P(Q|L)$ is large. But $P(S|Q)$ is more likely than $P(L|Q)$ because there are many more salesmen than librarians. An example of the Base Rate Fallacy. Evolutionary defect?

Students: Twenty Coin Flips?

- What is probability of exactly 10 heads?
- Of getting four heads in a row?
- Of getting all heads?

HTHHTHTTTHTHHTHTTHT

HHTTTHTHTTHTHTHTTTH

THHTTHTHTHHHTTHTHTH

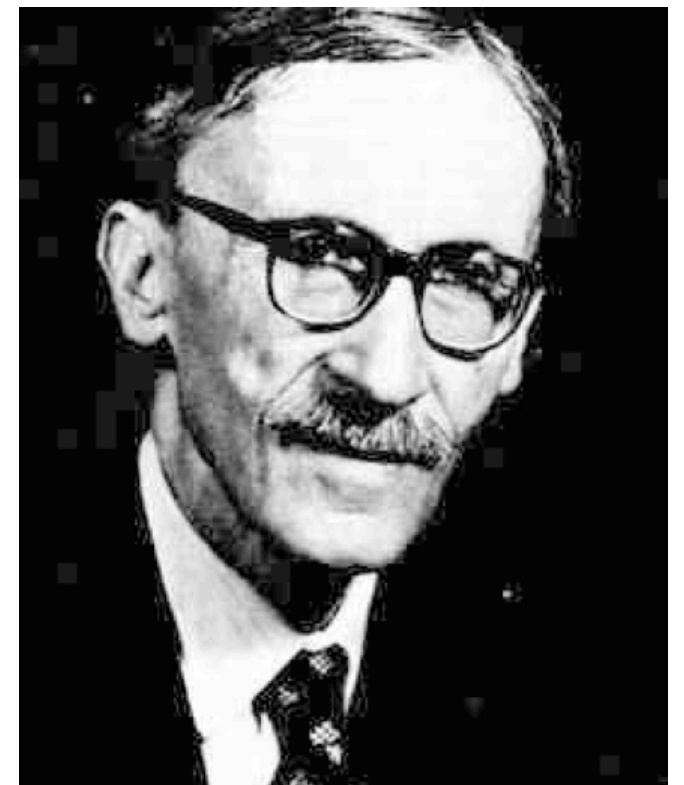
- Average fraction of heads as you flip more and more?
Does that define the probability for one flip? Try it.
How do you define probability for other events? E.g., rain?

Twenty Coin Flips?

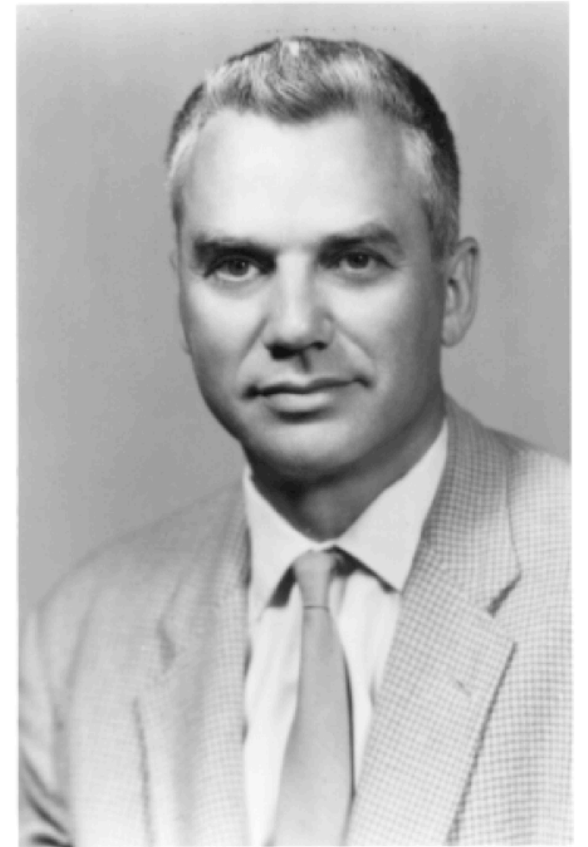
- What is the probability of exactly 10 heads? (.18)
- Of getting four heads in a row? (.77)
- Of getting all heads? (one in a million)
- Average fraction of heads as you flip more and more?
- Strong Law of Large Numbers (Borel 1909) says that the fraction of heads in a sequence of fair tosses tends to .5 except with vanishingly small probability.
- He avoided countable additivity! Didn't say "with prob 1."
- What do small probabilities mean? Crucial for linking mathematical probability with reality (Cournot).

And the Professionals?

- By my graduate days, the Kolmogoroff axioms were basis for the abstract approach everyone called “French Probability.”
- But there was originally much resistance there. E.g., Kolmogoroff’s contemporary, Paul Lévy.
- Loève called him “the great painter of probability.”
- Meyer writes of Lévy, “Despite his professorship...one often heard said that ‘he is not a mathematician’.”



- **About the great pioneer in this field, Doob wrote:**
“[Paul Lévy] is not a formalist. It is typical of his approach to mathematics that he defines the random variables of a stochastic process successively rather than postulating a measure space and a family of functions on it with stated properties, that he is not sympathetic with the delicate formalism that discriminates between the Markov and strong Markov properties, and that he rejects the idea that the axiom of choice is a separate axiom which need not be accepted. He has always travelled an independent path, partly because he found it painful to follow the ideas of others.”



- **Of Lévy’s seminal book, Paul Meyer wrote:**
“Like all of Lévy’s work, it is written in the style of explanation rather than proof, and rewriting it in the rigorous language of measure theory was an extremely fruitful exercise for the best probabilists of the time (Itô, Doob).”

- And Itô wrote: “During those five years [1938-43] I had much free time, thanks to the special consideration given me by the then Director Kawashima ... Accordingly, I was able to continue studying probability theory, by reading Kolmogoroff's *Basic Concepts of Probability Theory* and Lévy's *Theory of Sums of Independent Random Variables*. At that time, it was commonly believed that Lévy's works were extremely difficult, since that pioneer of the new mathematical field explained probability theory based on his intuition. *I attempted to describe Lévy's ideas, using precise logic that Kolmogoroff might use.*”

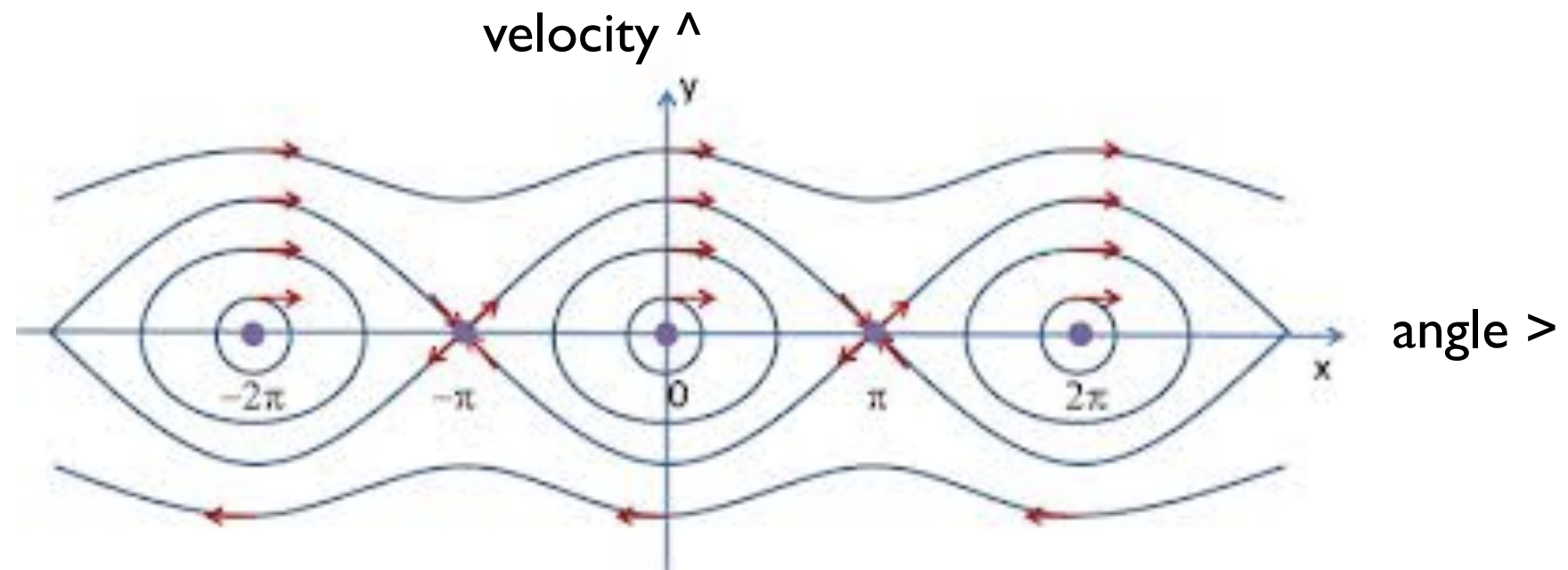


Mathematical Probability??

- Why work with such axioms?
- Most people don't respect them.
- Great mathematicians didn't respect them.
- The triple (Ω, \mathcal{F}, P) seems especially mysterious.
Any examples at all of what these are and how used?
- Let's see what Poincaré had to say about omega;
what Wiener had to say about measures;
and what Itô had to say about sigma-fields.

King Oscar's 60th Birthday Prize in 1889

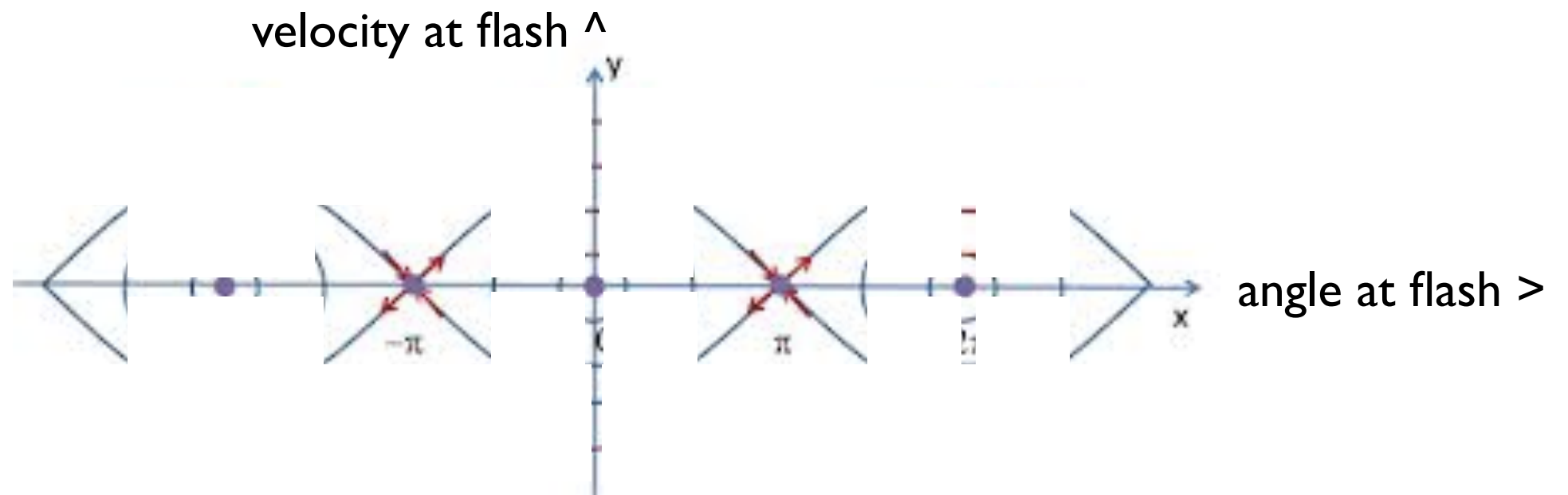
- One Prize Problem was the Stability of the Solar System.
- Model 3-bodies, restricted, as a periodically forced pendulum.
- Poincaré considered phase space of all solution curves.
- In unforced case here, note asymptotes to unstable equilibrium.



Uniqueness means that solution curves do not cross.

Stability of Forced Pendulum?

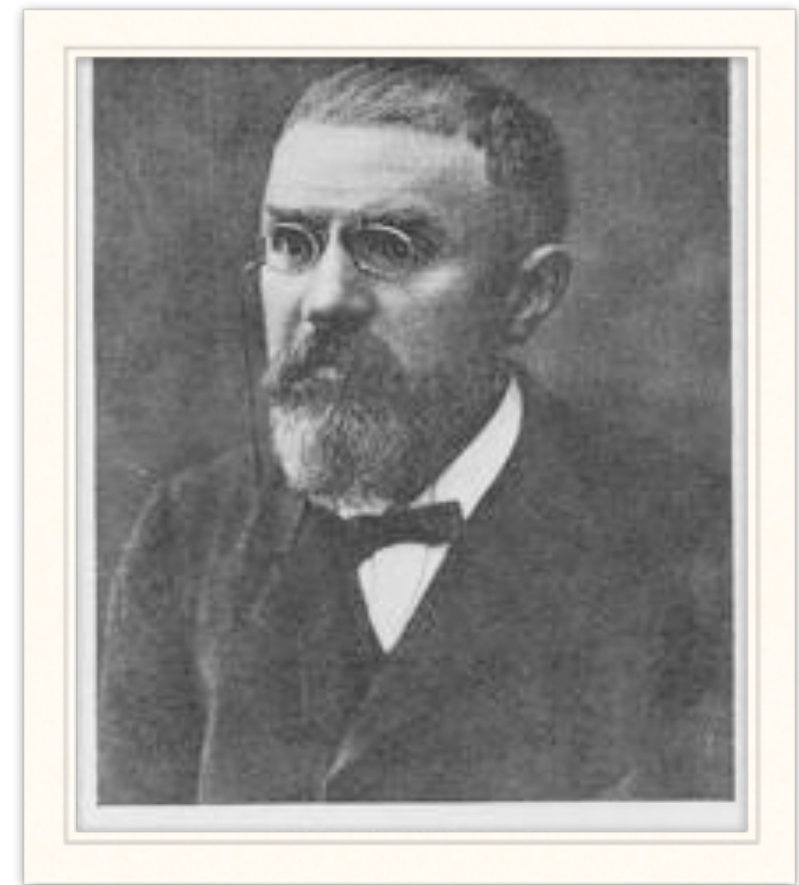
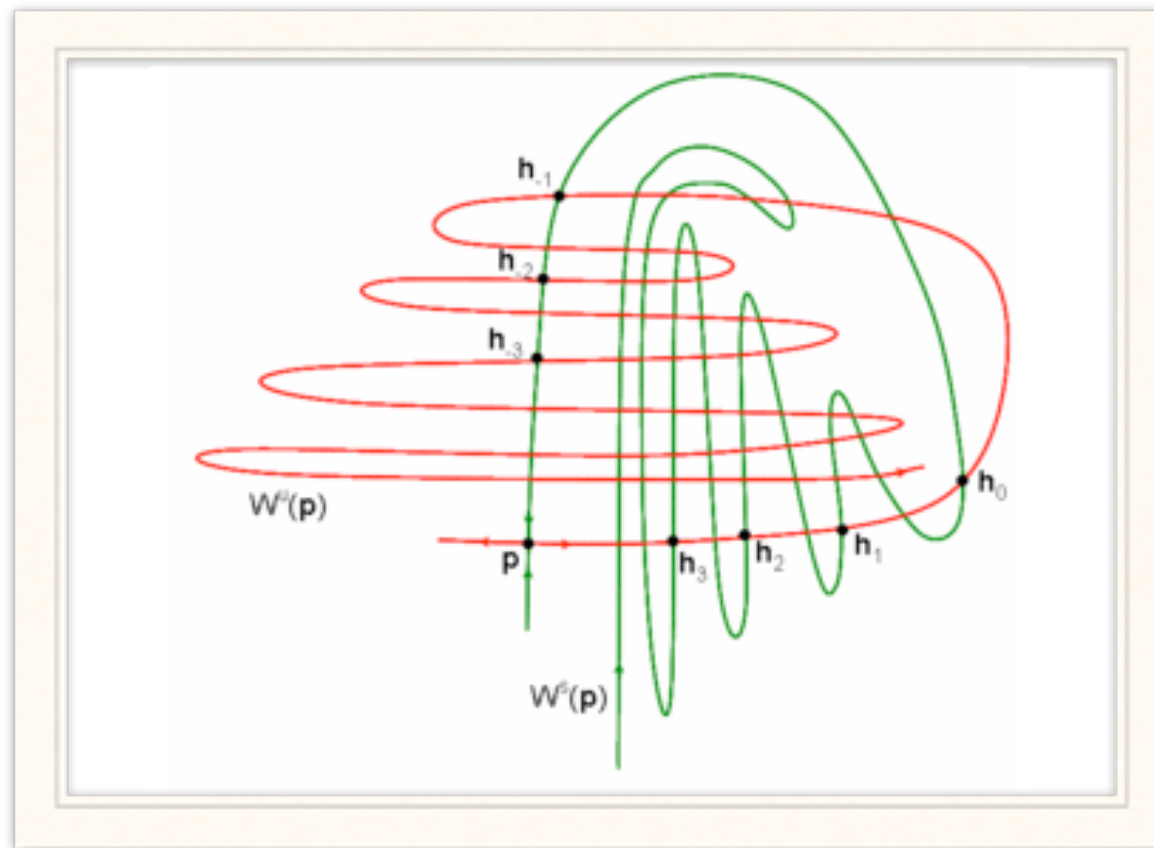
- Look at return map (stroboscopic). Jumps not flows.
- Stable and unstable equilibria (up and down) persist.
- But not all the periodic solutions that repeated over and over.
- Stable and unstable manifolds persist near unstable equilibrium.



Area preservation shows they can't avoid each other..

Poincaré's Tangle

- If those manifolds cross transversely at one point, they must cross again the image point and at the preimage point. Then their images and preimages have to cross, too.
- Would think this impossible, but editor asked about it.
- Implies chaos, but scandal suppressed this finding.

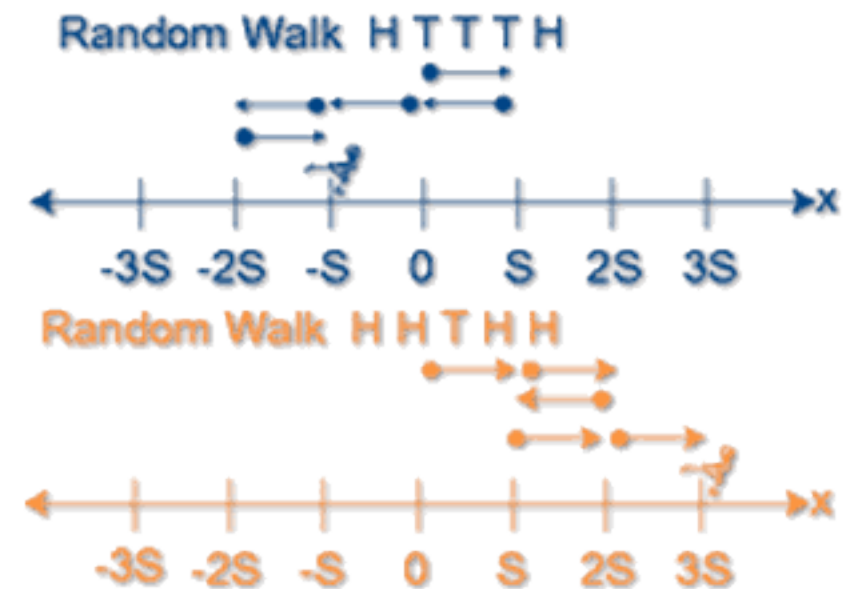
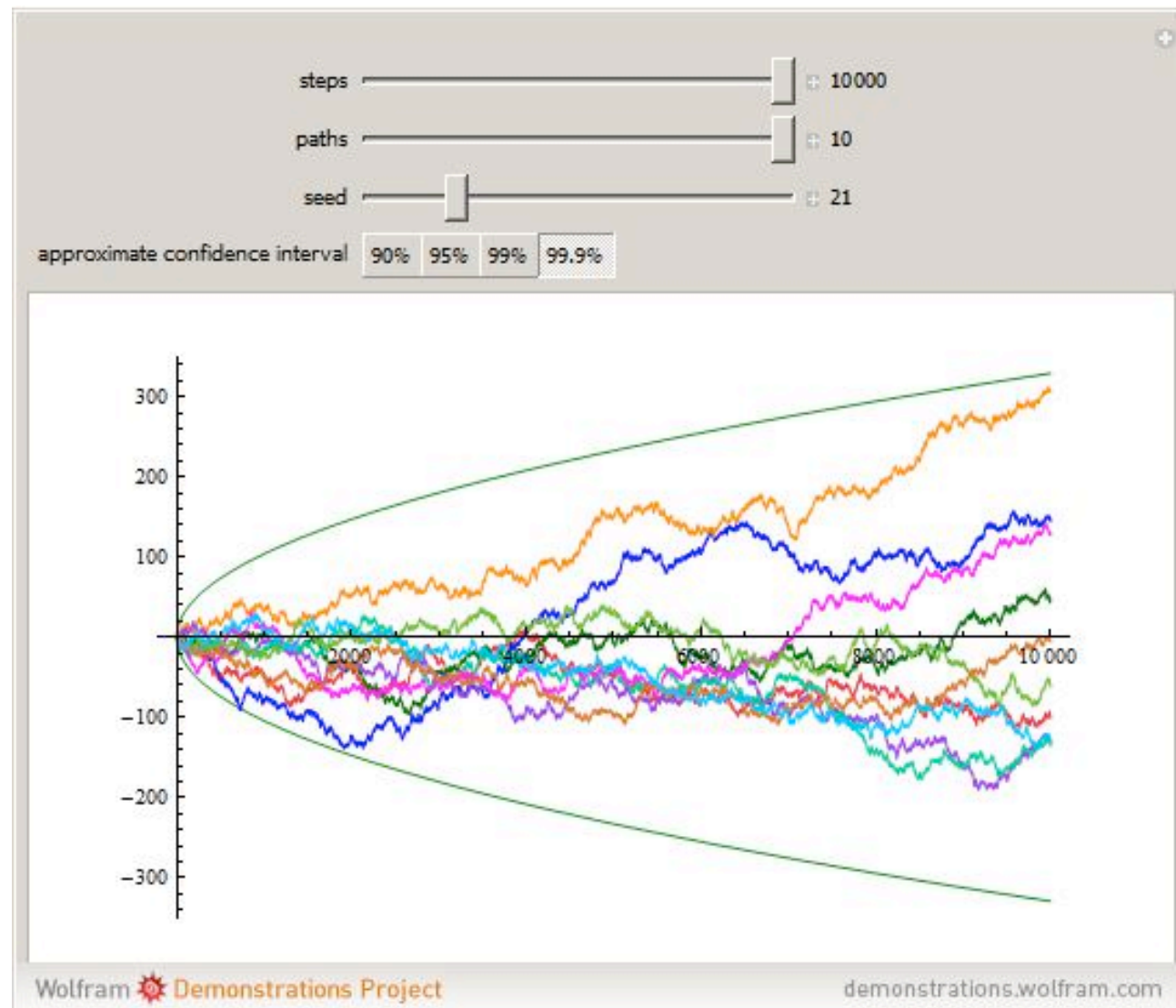


Chaos

- Graph of a function with unbounded variation.
- Smale later showed you get “horseshoes.”
- I.e., get theorems like this: write down R or L when pendulum passes through lowest point. Any sequence is possible and the dynamics in the phase space is conjugate by a measurable map with a Bernoulli Process (in other words, the deterministic system acts like coin flipping!).
- So Kolmogoroff model can, indeed, be found in nature.
- Note: Liouville measure is preserved in mechanics. Poincaré used this to prove recurrence of almost every trajectory on a bounded region of phase space.

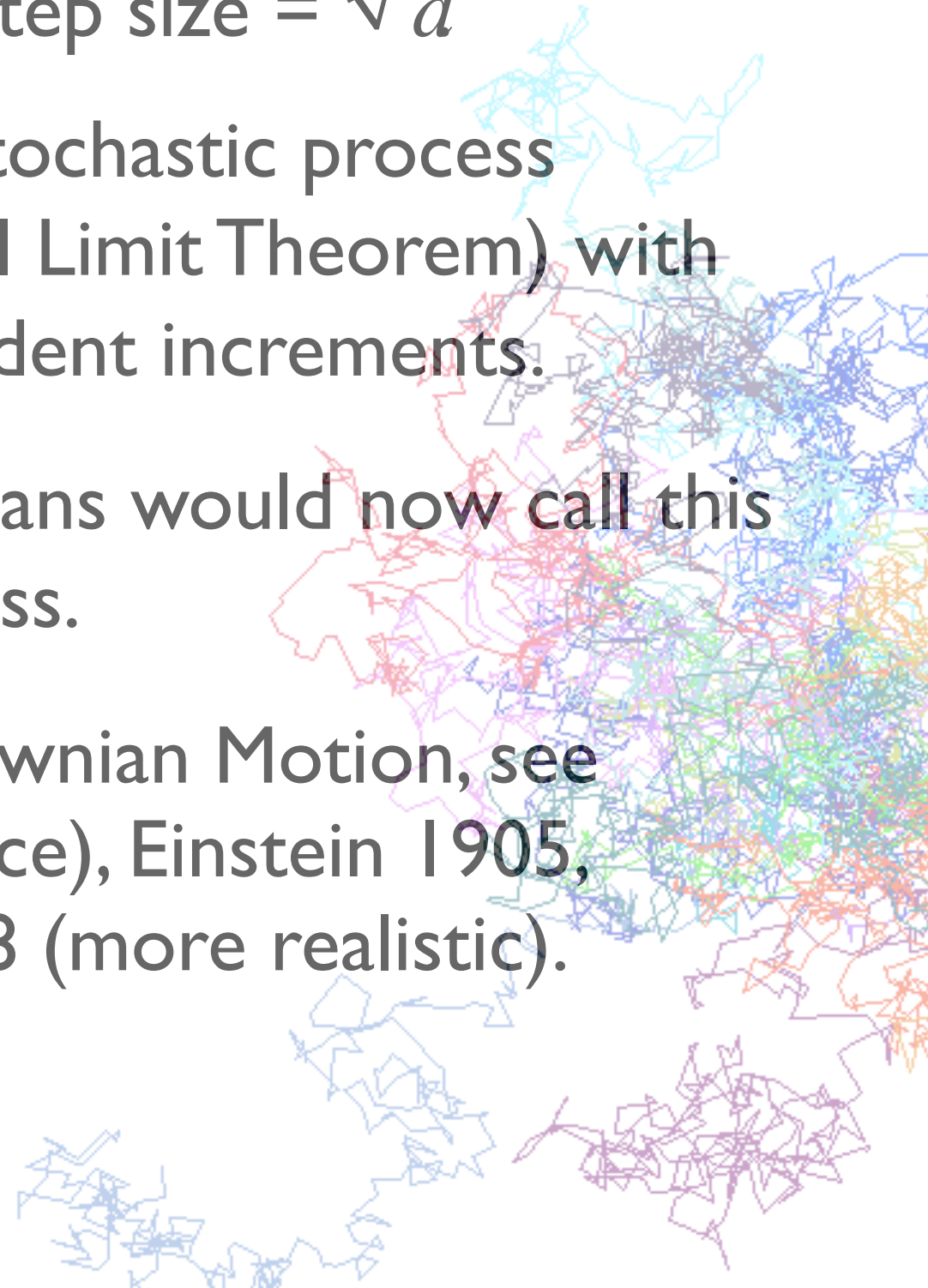
Add Up Flipping to Get Walking

- Suppose you step left or right based on coin flip.
- Get a stochastic process $X_n =$ your location after n steps.
- Called a random walk.



Scaling Limit of Random Walk

- Fix a real $t \geq 0$, then take the limit as the number of steps n goes to infinity with $t = nd$ and step size $= \sqrt{d}$
- As d goes to zero, seem to get a stochastic process $W(t)$ that is normal (by the Central Limit Theorem) with mean zero, variance t , and independent increments.
- If also continuous in t , mathematicians would now call this $\{W(t)\}$ the standard Wiener Process.
- In applications, others called it Brownian Motion, see Brown 1827, Bachelier 1900 (finance), Einstein 1905, Smoluchowski 1906, Langevin 1908 (more realistic).
- Like adding up lots of white noise.



Does Such a Process Even Exist?

- Wiener approached this several ways beginning in 1923.
- Showed that, with probability one, paths are of unbounded variation on every interval. (Wiggle wildly like the tangle.)
- Intuitively, paths nowhere differentiable since $W(t+h)-W(t)$ has variance h , so $(1/h)\{W(t+h)-W(t)\}$ has variance $1/h$.
- Requires a measure μ on space of continuous paths to make it a probability space. Now called Wiener measure.
- In fact, can then think of a Wiener Process as just a single function-valued random variable from the probability space (Ω, \mathcal{F}, P) to the probability space $(C[0, \infty), \Sigma, \mu)$ whose elements are continuous functions on $t \geq 0$.

Wiener Measure?

- Consider finitely determined cylinders $S \subset C[0, \infty)$ of form:

$$S = \{B \in C[0, \infty) : B_{t_j} \in A_j \text{ for } 1 \leq j \leq n\} \text{ where } 0 \leq t_1 < t_2 \leq \dots < t_n$$

and A_1, \dots, A_n are Borel sets in \mathbb{R} with product A .

- Using the Gaussian densities of a Wiener Process, we set:

$$\mu(S) = k_n \int_A \exp \left[-\frac{1}{2} \left\{ \frac{(x_n - x_{n-1})^2}{(t_n - t_{n-1})} + \dots + \frac{(x_1 - x_0)^2}{(t_1 - t_0)} \right\} \right] dx_1 dx_2 \dots dx_n$$

where $k_n = \left\{ \sqrt{2\pi(t_n - t_{n-1}) \dots (t_1 - t_0)} \right\}^{-1}$. The limit looks like:

$$d\mu = k \exp \left[-\frac{1}{2} \int_0^t \dot{x}^2(s) ds \right] Dx, \text{ which is suggestive nonsense.}$$

Homogeneous Chaos

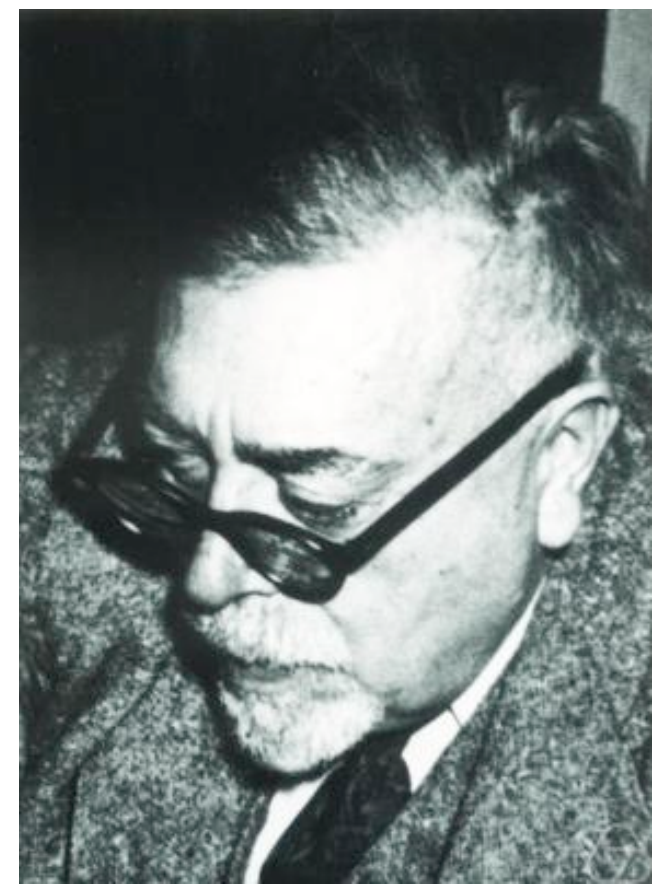
- So have to work much harder to obtain Wiener Measure.
- Assuming it exists, one way Wiener constructed Wiener Processes on $[0,1]$ was to take a sequence of Gaussian random variables Y_1, Y_2, Y_3, \dots with mean 0 and variance 1 defined on some probability space (Ω, \mathcal{F}, P) and some orthonormal basis $\{\varphi_n\}_{n=1,2,\dots}$ of the space $L^2[0,1]$ of square summable functions on $[0,1]$. Then the random sequence

$$W(t, \omega) = \sum_{n=1}^{\infty} Y_n(\omega) \int_0^t \varphi_n(s) ds$$

is uniformly convergent and so defines a continuous path with P probability one.

- Such an orthogonal basis for $L^2[\mu]$ is what Wiener called a homogeneous chaos. Gives a useful decomposition.

Eccentricity



- Many stories about Wiener.
- Existence of Wiener Measure and his Homogeneous Chaos not settled rigorously until 1968 by Itô and Nissio.
- Some therefore refer to the Wiener-Itô integral and the Wiener-Itô decomposition of $L^2[\mu]$.
- Itô was polite in his introduction to Wiener's papers. Speaking of subsequent work of Lévy, Kakutani, Doob, Kac, and himself, he wrote: "It is astonishing that all such developments stand on the basis given by Wiener's work on Brownian motion."

Itô Integral

- The challenge was to define the integral of a stochastic process along the path described by a Wiener Process, i.e.,

$$I(\omega) = \int_{s=a}^b X_s(\omega) dW_s(\omega)$$

- Wiener had handled deterministic integrands by integrating by parts and using ideas of Daniell.
- Usual approach would be to start with simple functions and a partition of $[a,b]$, then approximate like Riemann:

$$I \approx \sum_{i=0}^n X(s_i) [W(t_{i+1}) - W(t_i)] \text{ with } t_i \leq s_i \leq t_{i+1}$$

- Problem here is that, because the paths are of unbounded variation, different ways of choosing the s_i actually matter! Too much wiggling, even over shrinking subintervals.

Integrands Must Adapt

- Idea: deal with integrands that don't depend on the future.
- Filtration of (Ω, \mathcal{F}, P) is a family of sigma fields such that, for all $s < t$, we have $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$. Think \mathcal{F}_2 = events known at $t=2$, e.g., $A = \{\text{HHHH}, \text{HHTH}, \text{HHHT}, \text{HHTT}\}$ is “2 heads first.” Call different ω by same name. Reveals gradually, contra Lévy.
- Say that a process $\{X_t\}$ is adapted to the filtration generated by $\{W_t\}$ if each X_t is measurable with respect to that \mathcal{F}_t .
- Get independence using left end point for simple functions:

$$I \approx \sum_{i=0}^n X(t_i) [W(t_{i+1}) - W(t_i)]$$

- So for square summable X_t , this converges to define Itô's

$$I(\omega) = \int_{s=a}^b X_s(\omega) dW_s(\omega)$$

Itô's Calculus

- Itô Integration is just what is needed in many applications, especially in finance where you don't get to see ahead.
- Used this to solve Stochastic Differential Equations like:

$$dX = b(X, t) dt + \sigma(X, t) dW$$

- For a solution X , and $Y(t) = f(X(t), t)$, Itô's Formula says:

$$dY = \left[\frac{\partial f}{\partial t}(X, t) + b(X, t) \frac{\partial f}{\partial x}(X, t) + \frac{1}{2} \sigma^2(X, t) \frac{\partial^2 f}{\partial t^2}(X, t) \right] dt + \sigma(X, t) \frac{\partial f}{\partial x}(X, t) dW$$

- This is a strange chain rule. For $Y = f(X)$, $b = 0$, $\sigma = 1$, have:

$$dY = f'(X) dX + \frac{1}{2} f''(X) dt$$

Abstraction and Application

- Abstract from nature to math. Leave out lots. Then, the math forces surprising conclusions. Bigger surprise is that these, in turn, can tell you something new about nature if you are careful.
- Even though Weiner Process, say, moves infinite distances, has no velocity, and is otherwise unrealistic. Even if people, including coin flippers, don't behave ideally.
- Probability is also about people and how they bet. Bayesian: Subjective probability calibrated by frequency!
- Goethe: "Mathematicians are like Frenchmen."
- Itô was surprised to win prizes for applied math, too. His probability is part of mathematics now in any case.

Itô's Music

Only mathematicians can read "musical scores" containing many numerical formulae, and play that "music" in their hearts. Accordingly, I once believed that without numerical formulae, I could never communicate the sweet melody played in my heart. Stochastic differential equations, called "Itô Formulae," are currently in wide use for describing phenomena of random fluctuations over time. When I first set forth stochastic differential equations, however, my paper did not attract attention. It was over ten years after my paper that other mathematicians began reading my "musical scores" and playing my "music" with their "instruments." By developing my "original musical scores" into more elaborate "music," these researchers have contributed greatly to developing "Itô's Formula."

