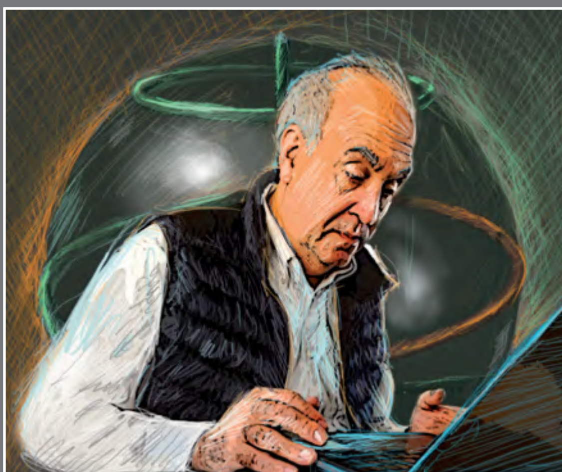
Daniel Goroff

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



*Prisme* N°37
October 2019

The Cournot Centre and Foundation

## The Cournot Centre and Foundation

The Cournot Centre is an independent, non-profit organization. It has created a forum where specialists meet to advance theories and increase understanding of the economic and social sciences and their epistemology. Like the Centre, the Foundation pursues catalysing work in the tradition of Augustin Cournot, accelerating theoretical formulations. Under the aegis of the Fondation de France, the Foundation puts into perspective the probabilistic paradigm shift originating in mathematics and spreading across disciplines.

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

**Daniel Goroff**

*Prisme* N°37

October 2019

Kiyosi Itô is a culminating hero of the story of how probability became a part of mathematics. It did not start out that way. This is a tale that includes a scandal involving Henri Poincaré, some eccentricity in the person of Norbert Wiener, and the music of Itô.

Although probability was not always part of mathematics, we must begin by trying to think of what *is* mathematics. There are many definitions. Let us start with Poincaré, who said, "Mathematics is the art of calling different things by the same name". This is a profound statement. It has a lot to do with abstraction, which was his definition of mathematics. Now, what did Poincaré think about probability? In 1912, he said, "One can scarcely give a satisfactory definition of probability." In his university years, Itô wrote that he doubted whether probability was an authentic mathematical field. Ironically, 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 very difficult to grasp, and people have known that for a long time. Let us look at Hilbert's Sixth Problem concerning mathematical physics:

> *6. Mathematical Treatment of the Axioms of Physics. The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which already today mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.*[1]

It thus specifically called for an axiomatization of probability. Why is probability so hard? Let us go back to Poincaré, who said in 1896, that when learning mathematics, it is a good idea to remember what the biologists say: ontogeny recapitulates the phylogeny. In other words, the development of the individual recapitulates, in some sense, the same stages that were gone through by the species. Think about the species and the development of geometry, for example: geometry has been around for thousands of years, so the species was able to do that rather quickly. Euclidean geometry is not easy, but many people do it even at the high-school level.

Probability took a very long time to develop into its own field. Two traditions have been built up on two questions: the question of equal probabilities and the question of small probabilities, also known as Cournot's principle,[2] which states that small probabilities do not really occur. Those are the two basic ideas, but nobody knew what to make of them, and they were the source of great controversy. Even the basic question – "What is a random variable?" – was difficult. One of my Harvard professors explained that students have a difficult time with these concepts, because no one ever defines for them what a variable is – not a random variable, which is even harder –, but simply a basic variable. Sometimes it is equal to 2, sometimes it is equal to 4.

Now let us turn to Bertrand Russell's definition of mathematics. In 1918, he said, "Mathematics may be defined as the subject in which we never know what we are talking about nor whether what we are saying is true".[3] People were indeed talking about these things, but they had no idea what they really meant.

This all changed for Itô when he was working at the Japanese National Statistical Bureau. He 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 Kolmogorov. Realizing that this was exactly what I had been looking for, I read through the book in one sitting. In* Grundbegriffe der Wahrsheinlichkeitsrechnung *(*Basic Concepts of Probability Theory*), written in German in 1933, Kolmogorov 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.* (Itô, 1998[4])

Andreï Kolmogorov's book was thus the key for Itô to defining probability as a field of mathematics. Other mathematicians had previously tackled this problem, but Kolmogorov is the one who completed probability's modern formalism, as Glenn Shafer and Vladimir Vovk remind us.[5] Borel, in 1905, noted that the new integration founded by Lebesgue on the concept of sets was an ideal tool for formulating probabilistic questions.[6] Fréchet, in 1915, extended the notion of the integral using the concept of abstract measure. Kolmogorov used their research to formulate a fully constructed model.

Kolmogorov defined a measurable space as a pair $(S, \Sigma)$, where $S$ is a set and $\Sigma$ is a sigma-field, that is, a collection of subsets of $S$ that includes $\emptyset$ and $S$, and that is closed under countable set operations. He defined a probability space as a triple $(\Omega, F, P)$, where $(\Omega, F)$ is a measurable space. We can think of $\Omega$ as the space of all possible worlds, present and future, associated with the space of all possible subsets of $\Omega$. As this last set is often considered to be too broad, we talk rather about the sigma-field and its elements, which are events $F$. Such events are often represented by Venn diagrams, such as the one in Figure 1. Basically, $\Omega$ is simply a place where we draw the Venn diagrams.

In Figure 1, the circle on the right represents the event of rain. We group together a lot of different worlds in one set, which represents the event that "*it is raining*". We can also say, *it is raining* and *it is Tuesday*. The idea that "it is Tuesday" is another event, represented in the circle on the left. You can intersect them, take their unions. The definition of a sigma-field indicates that you can do this for sequences of sets. You can do everything with set operations – unions, compliments, intersections – and still remain in that sigma-field of events. That is an important point, because if we try to take arbitrary subsets of a space, we get into trouble with trying to assign a measure in a sensible way. There are many kinds of paradoxes that one can fall into, so we must restrict ourselves to some sort of sigma-field.

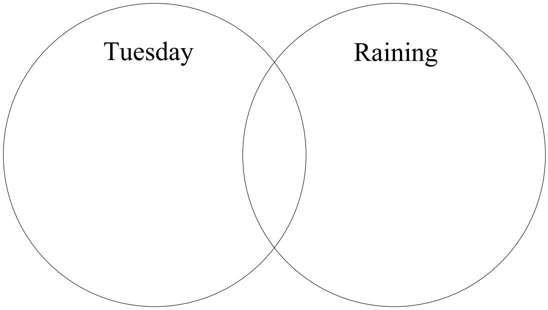



*Figure 1: Venn diagram*

This was a great innovation, because it made it possible to have a measure that is defined on the sigma-field in a consistent way. If it is a measure that is countably additive, in other words, if we have a series of disjoint sets, the measure of that whole sequence will be the sum of the measures of the sets individually. If the measure of the entire space is one, then we can call that a probability measure $P$, that is, a non-negative (between 0 and 1) and a countably additive set function on the measurable space with total mass 1. We now take that for granted, but this was not obvious before.

Continuing with Kolmogorov's probability axiom, $A \in F$ are called measurable sets, which are thought of as events. The measure assigns to each a number $P(A)$ between 0 and 1 that we interpret as the probability of that event. Note, for example, that

$$A \subset B \Rightarrow P(A) \leq P(B).$$

Before Kolmogorov, a lot of work had been done on measures, on countably additive set functions, but it was not called "probability". Other work was done on probability, but there was confusion as to whether things were countably additive. Many questions remained, such as: how could a point have no mass, but when points are put on a line, then the line has mass? Everyone had trouble grasping these ideas. There are, nevertheless, very simple intuitions that have come out of this. For example, if we have an event and a sub-event, the probability of the event must be bigger than that of the sub-event, because if we consider the compliment of the two, with the inner one, it is disjoined. There was really nothing else to probability, except for that sort of additivity.

Now let us turn to random variables. A random variable is simply a measurable function on a probability space. To be measurable, the function can only depend on the events. It has to be constant on any of these events, because it only depends, for example, on whether it is raining, or whether it is Tuesday. It does not depend on whether I am in this world or that world (outside the sigma-field). There are a lot of things that we do not distinguish; we call them by the same name by putting them into one of these elements of the sigma-field.

A random variable is a measurable function $X$ from a probability space ($\Omega$, $F$, $P$) to a measure space ($S$, $\Sigma$) 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.

Kolmogorov defined expectation as an integral:

$$E(X)=\int X(\omega)\,dP(\omega),$$

and conditional probability, given an event $B$, such that $P(B) > 0$, by the following expression:

$$P(A|B)=\frac{P(A\cap B)}{P(B)}.$$

Furthermore, Kolmogorov used the arguments of measure theory to extend the notion of conditional expectation, given an event $B$, and defined the conditional expectation, given a random variable.[7]

Now, let us look at what Joseph Doob said in 1997:

> *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, 2pi] when endowed with Lebesgue measure divided by 2pi 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.* (Doob, 1997[8])

The axioms were thus a rhetorical contribution, not research. The important thing was to write them down. People knew about measures beforehand; they even knew about measures that assigned 1 to the whole space. The amazing achievement was to be able to call it probability. Before that, people talked about these two things as if they were quite different. This was thus a rhetorical kind of triumph. Itô needed to work through these things and define them in order to use them to study stochastic processes.

As we understand it these days, a stochastic process is simply a collection of random variables indexed by a set (usually time). We can think of it as a series of successive coin flips: $\{X_1, X_2, X_3, \ldots\}$. The sigma-field is generated by finite subsets of those random variables that we assign something to in order to make them measurable. If I say that on flip numbers 3, 7 and 15, I get a head, a tail and a head, that is called a cylinder set. It is only a finite number of things: I know that the probability of each flip is 1/2; therefore, the probability of that entire set is 1/8.

Kolmogorov had a theorem that we can extend from the cylinders under many circumstances, and we get a sigma-field on the whole set. That way we skirt all of the questions on the probability of an infinite sequence, and everything else. It is possible, but it has to have probability 0, and so on.

We can think about the probability space just for these coin flips as all of the sequences of Hs and Ts. We have a shift map that shifts along and is called a Bernoulli Process. That is the simplest sort of example of a stochastic process. The point of this model is to have a random variable that is defined on this triple ($\Omega$, $F$, $P$) and takes values in the space of all Hs and Ts. So the $n^{th}$ flip is a measurable function:

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

Imagine Tyche, the Greek goddess of chance, reaching into the set of all these worlds and pulling out $\omega \in \Omega$, and that determines $X_n(\omega)$.

Now, if we think of that in terms of coin flipping, then let me recall what Johann Wolfgang von Goethe said, “Mathematicians are like Frenchmen: whatever you say, they translate into their own language, and forthwith, it is something entirely different.” So this is what a mathematician thinks of a coin flip; many other people think of a coin flip differently. It is actually not particularly natural or intuitive.

To drive home that point, I would like to explain how behavioral economists think of these simple ideas of probability. My example comes from Amos Tversky and Daniel Kahneman (1974).[9] It is an exercise that we do with students. We tell them that Linda is 31 years old. She is single, outspoken and bright. In college, she majored in philosophy and was concerned with discrimination, social justice and anti-nuclear rallies. Rank these probabilities from 1, the most likely, to 5, the least likely.

- 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.

Which do you think is most likely? And which do you think is least likely? Put them in the order that you think, given what you know about Linda.

Why am I asking these questions? When you ask people this question, the issue is who ranks “she is a bank teller *and* in the feminist movement” higher than the possibility that “she is a bank teller”. Most people do that. What is wrong with that? The problem is *AND*. Let us go back and think about the events. The event that she is a bank teller is large; the event that she is a bank teller *AND* active in the feminist movement is much smaller. This example shows that the very first and only principle of probability is violated routinely by people when you ask them questions of this kind. Consequently, this is not as natural and intuitive as we sometimes pretend.

Let me give you another example: In the first five pages of the typical English-language novel, how many six-letter words would you expect to find with the penultimate letter “n”, that is, of the form: - - - - n - ?

This is a probability question in another form. In five pages of a book, there are probably 3000 words. The question is, “What is the probability of finding this type of word?” What are the types of numbers that you would get? You would not say 1000. People say 1, 2, 3… 50, 60… We ask half of the class this question.

Then we ask the other half of the class the following question: In the first five pages of a typical English-language novel, how many six-letter words would you expect to find ending in “ing”, of the form: - - - ing?

To that question, the students’ answers are not 1, 2, 3, or even 50 or 60. People suppose many more, perhaps a few hundred words. This example shows again that estimating these sorts of probabilities greatly depends on how we ask these questions. There is, once again, a violation of the first and only principle of probability, which our minds tend to get mixed up about.

Let me give you another example: You know that Tom is either a salesman or a librarian. His personality has been described as quiet. Which is more likely, S or L?

Another question: Fred is either a salesman or librarian. You know nothing else. Which is more likely? Some people say one-half, because there are only two possibilities. We know that is wrong. There are a lot more salesmen in the world than there are librarians. In the United States, according to the Bureau of Labor Statistics, there are about 100 times more salesmen than librarians. That means that if even one in 10 salesmen happens to be quiet, it is more likely that you are looking at a salesman who happens to be quiet than a librarian. This is about conditional probability. We said we all understand Kolmogorov’s definition of probability, but it is not that simple. Even though the probability of a person being quiet when s/he is a librarian is high, that was not the question. The question was: For a person who is quiet, what is the probability of being a librarian, or a salesman? This is an example of the base rate fallacy: if presented with related base-rate information (that is, generic, general information) and specific information (information only pertaining to a specific case), the mind tends to ignore the former and focus on the latter. These are things that evolution has not made us good at. Nature, for its own reasons, has made this sort of axiom difficult.

To go back to the stochastic processes and coin flips, here is another question that you can ask students: tell them to go home, not flip a coin, but just write down Hs and Ts that could have come from flipping a coin. Sometimes I ask them to do 100, which takes a little time. But if you ask them to do only 20, what is the probability that the students' papers will have exactly 10 heads and 10 tails? Almost all of them do. If they can count, they will give you a sequence of 20 heads and tails with exactly 10 Hs and 10 Ts. What is the probability of getting four heads in a row from the students? Almost never. You never see all heads either.

You see something more like the following example:

HTHHTHTTTHTHHTHTTHT

HHTTTHTHTTHHTHTHTTH

THHTTHTHTHHHTTHTHTH

You can ask them about the average fraction of heads as you flip more and more. If it is a fair coin, everyone believes that is supposed to be one-half. We think that is the definition of the probability of one flip, when you can do things over and over again. To test that correlation between abstractions and the real world, a researcher I know conducted an experiment in which he filled a high-school gymnasium with students and asked them to flip coins for several hours. One student flipped; the other wrote down "heads" or "tails". This experiment gave him huge amounts of data about coin flips. Nevertheless, when he applied statistical tests to verify the data, he failed all of the tests! The data was terrible and did not look independent; everything came out about half-and-half. He could not understand what had gone wrong. When he went back and checked the experiment, he realized that at the beginning of the experiment, the students flipped the coin high up in the air and wrote down the result. After about 10 minutes, however, the students got tired, and instead of flipping the coin high enough for it to turn over several times, they only flipped it a little bit, and the coin only turned over two or three times. So once they got tired, autocorrelation occurred, and all the statistics went wrong. This example shows that when we try these sorts of experiments, we get a sequence, and we think that is what the probability is, but it does not always work in nature.

Assigning probability to cases that you cannot repeat is another interesting question; this is particularly the case in decision theory. Sometimes I ask students, “What does it mean to say that the probability of rain tomorrow is 80%”? They answer, “If you had 10 days, it would rain eight of those days”. Does that make sense? No, that does not make sense. If you had 100 days just like today, it would rain 80 of those days? None of that makes sense, because you never have 100 days just like today.

Continuing with the coin flips and stochastic processes, the probability of getting exactly 10 heads is less than 0.2, even though all the students write it down. The probability of getting four heads in a row, when you write down 20, is almost 80 per cent (even though they never write it down). The probability of getting all heads is about 1 in 1 million. For the average fractions, there is the strong law of large numbers, originating from the work of Emile Borel. In his fundamental paper of 1909, Borel defined the notion of convergence with probability 1 and formulated a first statement of the strong law of large numbers.[10] He said that the fraction of heads in a sequence of fair tosses tends to 0.5, except with a vanishingly small probability.

This is where the issue of small probabilities comes in. If we want to find the probability in the way I just described, we do a sequence, take the average, and if it comes out to 0.5 that must be the probability for one coin flip. That is a problem, because if it only happens with probability 1, how do we know that this time, it worked? You need this interpretation to make probability theory join with the real world; you need the Cournot principle. People have argued back-and-forth about it. That is another reason why the Kolmogorov axioms made such a difference in helping people to think about these concepts.[11]

Let us now turn to how the professionals reacted when this axiomization came out. When I was a graduate student, people talked about “French probability”. Ironically, there was a lot of resistance to probability in France, and, in particular, to the ideas of Paul Lévy.[12]

Lévy began his career as an analyst, specializing in functional calculus. After WWI and after publishing the works of René Gateaux, he turned to probability and observed an extraordinary connection between functional calculus and probability.[13] Jacques Hadamard, his Ph.D. adviser, had hoped that Lévy would be the great future specialist of functional calculus and regretted his change of discipline.

Again, we come up against the question as to whether probability was considered to be part of mathematics. Whatever the case, Lévy carried probability further into the discipline (Barbut et al., 2014). Doob wrote the following about Lévy, as a great pioneer in the field of probability:

> *[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 traveled an independent path, partly because he found it painful to follow the ideas of others.*[14]

Paul-André Meyer wrote of Lévy's 1948 seminal book on stochastic processes and Brownian motion:[15] "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)" (Meyer, 2009 [2000]).[16] It was indeed Itô's vision that took Lévy's intuitions and beautiful painting and changed it into something rigorous and more like mathematics.

Itô himself said,

> *During those five years [at the Cabinet Statistics Bureau, 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."* (O'Connor and Robertson, 2001[17])

Coming back to the idea of mathematical probability, the question arises of why should we work with these axioms? We have already seen that most people, including a lot of great mathematicians, do not respect them. The triple ($\Omega$, $F$, $P$) remains very mysterious. Are there any examples of how it is used and why it comes up? Everybody does it now, but why?

To answer those questions, let us look at Poincaré's position on $\Omega$, Wiener's interpretation of measures and Itô's take on sigma-fields. In 1889, King Oscar II of Sweden celebrated his 60th birthday. Mittag Leffler organized a prize for offering a collection of mathematical works to the King. One of the proposed problems was to show something about the stability of the solar system. Poincaré chose that topic and modeled the restricted three-body problem as a periodically forced pendulum.[18] We can consider the space of all solution curves, much like $\Omega$: we are looking at all of the possible solutions. Poincaré was the first one to start drawing pictures like this:

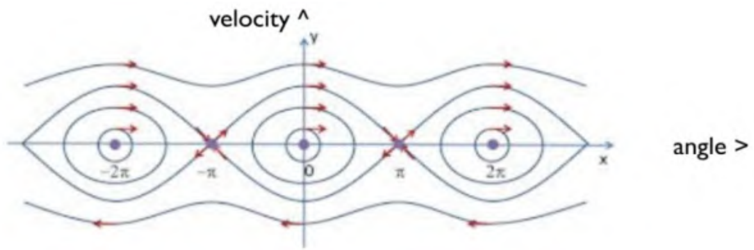



*Figure 2: Unstable equilibrium of the three-body problem*

He had the idea that if we periodically force it, then we can see a stroboscopic picture at every period. If we write that down, we get some of these features. He assumed, however, when he was doing it, that the asymptotic curves that came out of the unstable equilibrium – the straight-up equilibrium – would have to connect nicely with one another. He thought he had proven the stability results for the solar system. He won the prize, but then one of the editors said, "How do you know that they don't cross instead of joining and making a nice stability result"?

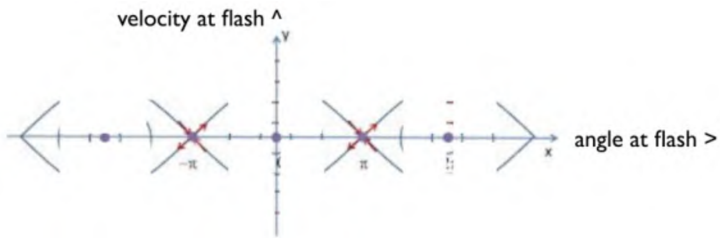



*Figure 3: Phase space at flash*

Poincaré thought about it and eventually realized that he was completely wrong. He had not proven the stability result, because if those manifolds cross transversely at one point, they must cross again at the image point and at the pre-image point, which means that their images and pre-images have to cross too (Figure 4).

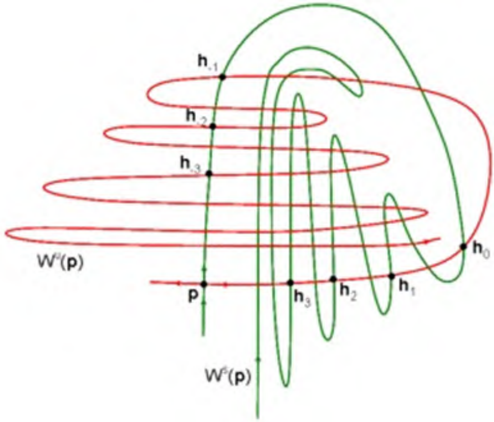



*Figure 4: Poincaré tangle*

This is now called the Poincaré tangle. From all these asymptotic curves, we get a big mess, and what results from that mess is now called “horseshoes”.

We obtain chaos theorems like the following. Let’s say that we have a forced pendulum, and we write down the results. The pendulum makes some crazy motions, but every time it passes through the lowest point, we write down an R if it comes from the right or an L if it comes from the left. This is a way of coding those crazy motions. The theorem that we can prove is that any sequence is possible. We thus obtain a space of all the sequences of Rs and Ls. This is completely deterministic. For the correct initial conditions, we could have it go R, R, R, R, R from the beginning of time, then all of a sudden, tomorrow, it goes to L. It could do three Rs, one L, four Rs. It is completely unpredictable. This should remind us immediately of coin flips: any sequence is possible.

This is a model of the Kolmogorov space, which can actually be found in the mechanics of a deterministic system – something that has a measure. It is called the “Liouville measure”, because mechanical systems have a preserve measure. It acts just like the coin flipping.

People got very excited, saying that we had found chaos in deterministic systems, which is indeed worth getting excited about.

What I find significant, but which has been mentioned very little, is that this is actually an example of the model for coin flipping. I am not sure that anyone has ever before produced a system in which we know what the omega is, what the function is, that we know what we are writing down, and we obtain a sequence of Rs and Ls. If we change those into Hs and Ts, we have something that is behaving the way a stochastic process is supposed to.

This is a case where, instead of being mysterious (and this is about the only case that I know of), we actually see that the omega in this construction makes sense; we obtain a stochastic process out of something that exists in nature and the mechanics of which everyone can understand and accept.

Now let us move from coin flipping to walking.

Suppose that we step right or left, based on a coin flip, then we get a stochastic process where $X_n$ is your location after $n$ steps, which is called a "random walk".

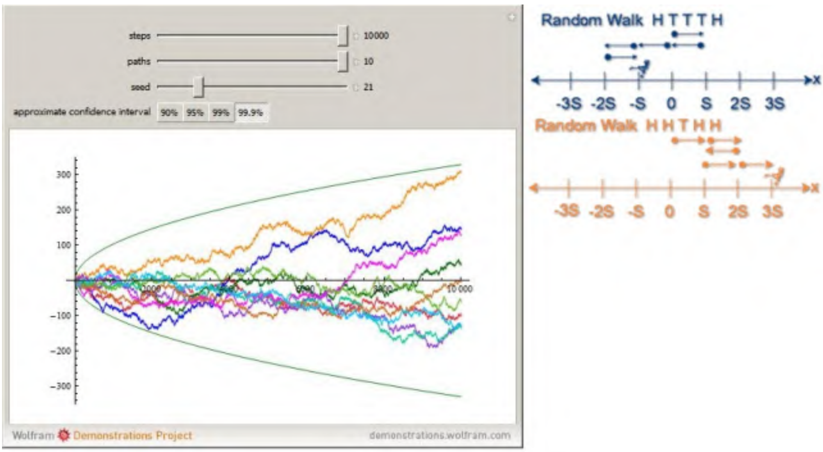



*Figure 5: Simulation of several independent random walks*

Here is what happens when we take a scaling limit of the random walk. First, we fix a real $t \geq 0$, and then take the limit as the number of steps $n$ goes to infinity with $t = n\delta$ and the step size is equal to $\sqrt{\delta}$. As $\delta$ goes to 0, we 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 it is also continuous in $t$, mathematicians would now call this $\{W(t)\}$, the standard Wiener Process.

This is what the paths look like:

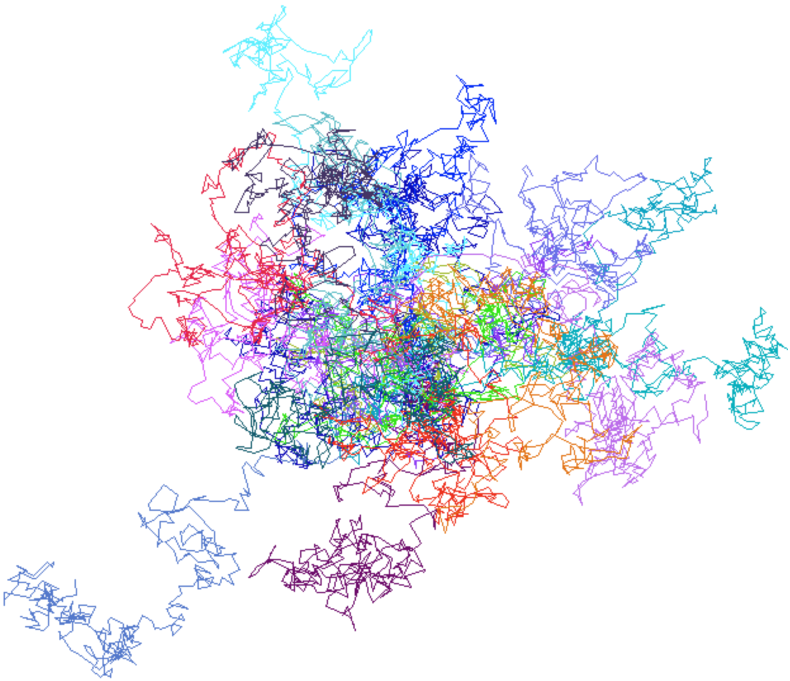


*Figure 6: Simulation of several independent Wiener processes*

In other applications, this was called Brownian motion for a long time (see Brown, 1828).[19] Bachelier (1900)[20] was the first to write about it in finance. Although they found his work interesting and original, the mathematicians of the time were not ready to consider finance as a worthy subject for applied mathematics, unlike physics, for example.

Bachelier did not arouse interest among them. He did, however, immediately catch the attention of actuaries.

In physics, Einstein (1905)[21] proposed the first model for Brownian motion, which was later extended by Langevin (1908),[22] before Marian Smoluchowski (1916) added his contribution.[23]

Brownian motion is a fascinating example of a highly vibrating motion. It is similar to adding up a lot of white noise rather than adding up a lot of steps in one direction or another.

There is a question as to whether Brownian motion even exists. It is a rather strange question, because Brown looked through the microscope and saw things moving around. People thought that they knew what Brownian motion was, and they had made a lot of abstractions and written them down. Then Einstein solved diffusion equations, making it possible to calculate Avogadro's number on the space. There was thus a physical sense that all this existed. Making a mathematical existence proof turned out to be very hard, however.

To build a mathematical model of Brownian motion, in 1923, Wiener used the principles he found in Lévy's first book, *Leçons d'analyse fonctionelle*.[24] One of the things that Wiener did was to investigate the properties of these paths to show that with probability 1, the paths are of unbounded variation on every interval. That means that they wiggle wildly like the tangle and are nowhere differentiable. Intuitively, paths are not differentiable since $W(t+h)-W(t)$ has variance $h$,

so $\left(\frac{1}{h}\right)\{W(t+h)-W(t)\}$ has a variance of $\frac{1}{h}$.

As $h$ goes to 0, it means the variance is huge, which is just another way of saying that it is very, very wiggly. To say that it blows up like that with probability 1 means that it requires a measure μ on the space of continuous paths to make it a probability space. That is now called the Wiener measure. We can, in fact, think of a Wiener Process as simply a single function-valued random variable from the probability space $(\Omega, F, P)$ to the probability space $(C[0,\infty),\Sigma,\mu)$ whose elements are continuous functions on $t \geq 0$.

If we want to make a Wiener measure, then, similar to coin flipping, we consider finitely determined cylinders $S\subset C[0,\infty)$ of form:

$S=\{B\in C[0,\infty): B_{t_j}\in A_j \text{ for } 1\leq j\leq n\}$ where $0=t_0<t_1<t_2...<t_n$.

And $A_1, \ldots, A_n$ are Borel sets in $\mathbb{R}$ with product $A$.

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

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

where

$$k_n=\left\{\sqrt{(2\pi)^n(t_n-t_{n-1})\ldots(t_1-t_0)}\right\}^{-1}.$$

The general expression would be:

$$d\mu=k\exp\left[-\frac{1}{2}\int_0^t x^2(s)\,ds\right]Dx,$$

which is suggestive, but does not have any obvious meaning.

If we have a finitely determined cylinder, the expression above $dx_1, dx_2,\ldots dx_n$, makes sense. But in the third line, we have an integral of a derivative of this path. As already mentioned, however, there is no derivative anywhere. This is, therefore, complete nonsense. Then we have some sort of flat measure on the space of all these, and that does not exist either. All this shows that when we do the obvious thing, we get into deep trouble.

People tried a lot of different ways of doing this. Wiener himself found a basis for $L_2[\mu]$.

Assuming it exists, one way Wiener constructed Wiener Processes on [0,1] was to take a sequence of Gaussian random variables $Y_1,Y_2,Y_3,\ldots$ with mean 0 and variance 1 defined on some probability space $(\Omega, F, P)$ and some orthonormal basis $\{\varphi_n\}_{n=1,2\ldots}$ of the space $L_2[\mu]$ 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 thus defines a continuous path with $P$ probability one. Such an orthogonal basis for $L_2[\mu]$ is what Wiener called "homogeneous chaos". He found some ways of saying that we could produce these sorts of paths, the Wiener paths. The truth, however, is that it was not settled rigorously until 1968, in a paper that Itô co-authored with Makiko Nisio[25] giving origin to the references of the Wiener–Itô integral and the Wiener–Itô decomposition of $L_2[\mu]$.

There are lots of stories about Wiener and his eccentricity. It is not surprising that it was very hard to make sense out of what he was doing: half of it was nonsense and half of it was brilliant. Itô was polite in his introduction to Wiener's papers. Speaking of the subsequent work of Lévy, Shizuo Kakutani, Doob and himself, he wrote, "It is astonishing that all such developments stand on the basis given by Wiener's work on Brownian motion".

Itô's challenge was to define the integral of a stochastic process along the path described by a Wiener Process. It was not the first time that an attempt was made to define a stochastic integral, but Itô's formulation proved particularly operational. It makes it possible to formulate a practical differential calculus, extending ordinary differential calculus:

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

Wiener had handled deterministic integrals by integrating by parts and using the ideas of Percy Daniell (the Daniell integral[26]). The usual approach would be to start with simple functions and a partition of $|a,b]$, then approximate, like Bernhard Riemann (the Riemann integral), by adding up the differences of the Wiener processes and weighting it by the stochastic process $X$ at some point in between the points on the partition:

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

The problem is that because the paths are of unbounded variation, different ways of choosing the $s_i$ matter. There is massive wiggling, and the answer that we get when we try this naive approach depends on how we chose the $s_i$ and the intervals of the partition. We can get almost any answer we want by doing this – even the over-shrinking of subintervals – just because there is too much wiggling.

Itô realized, therefore, that we must deal with integrals that do not depend on the future. This makes a lot of sense for finance, because the future is never known. The way of formalizing this was to think about a filtration. A filtration of $(\Omega, F, P)$ is a family of sigma fields such that, for all $s < t$, we have $F_s \subset F_t \subset F$. $F_2 = \{$events known at $t = 2\}$, for example: $A = \{$HHHH, HHTH, HHHT, HHTT$\}$ is "2 heads first". So, we say that a process $\{X_t\}$ is adapted to the filtration generated by $\{W_t\}$, if each $X_t$ is measurable with respect to $F_t$. So it only depends on the things known up to that time; it does not know anything about the future and cannot depend on it.

Continuing, we get independence using the left end point for simple functions:

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

This shows that you can take such a procedure and actually define an integral.

So for square summable $X_t$, this converges to define Itô's integral:

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

The Itô integral was the insight of thinking about these filtrations and defining them in such a way as to allow us to do Itô calculus. There are some wonderful formulae for Itô calculus. Itô Integration is just what is needed in many applications, especially in finance where you don't get to see ahead. Itô used this to solve Stochastic Differential Equations, such as:

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

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

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

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

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

To conclude, we abstract a lot from nature to mathematics, and we leave a lot of things out. Then the mathematics forces surprising conclusions. The bigger surprise is that these, in turn, can tell you something new about nature if we are careful. This is also true for Wiener processes: they move infinite distances, do not have a velocity and are otherwise unrealistic. People are not very realistic either when they flip coins, because they get tired and they do not do what the models have specified. Probability is also about people and how they bet. It is Bayesian: subjective probability calibrated by frequency! What Goethe said about mathematicians being like Frenchmen is important. The amazing thing about all of this is that Itô himself was very surprised when he won the Gauss prize, because it was for applied mathematics. He thought of himself by that point as a real mathematician. He also won awards for regular mathematics. And his probability is certainly part of mathematics now.

I would like to end by saying that he wrote a lovely little text about 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".*[27]

Itô made a tremendous contribution to ensuring that probability is a part of mathematics. I am particularly pleased to be able to honor him in this text.

## Endnotes

1. Hilbert, David (1902), “Mathematical Problems”, *Bulletin of the American Mathematical Society*, 8(10), pp. 437–479.

2. Touffut, Jean-Philippe (ed.) (2007), *Augustin Cournot: Modelling Economics*, Cheltenham, UK, and Northampton, MA, USA: Edward Elgar Publishing.

3. Russell, Bertrand (1918), *Mysticism and Logic and Other Essays*, London: George Allen & Unwin LTD., 3rd edition, p. 75.

4. Itô, Kiyosi (1998), *My Sixty Years in Studies of Probability Theory: Acceptance Speech of the Kyoto Prize in Basic Sciences*.

5. Shafer, Glenn, and Vladimir Vovk (2006), “The Sources of Kolmogorov’s Grundbegriffe”, *Statist. Sci.*, 21(1) , pp. 70–98.

6. Durand, Antonin, and Laurent Mazliak (2011), “Revisiting the Sources of Borel’s Interest in Probability: Continued Fractions, Social Involvement, Volterra’s Prolusione”, *Centaurus*, 53(4), pp. 306–332.

7. Charpentier, Eric, Annick Lesne and Nikolaï K. Nikolski (eds) (2007), *Kolmogorov's Heritage in Mathematics*, Berlin, Heidelberg: Springer-Verlag.

8. Snell, J. Laurie, “A Conversation with Joe Doob”, https://www.dartmouth.edu/~chance/Doob/conversation.html.
This interview provided the basis for an article in *Statistical Science*, 12(4), November 1997.

9. Tversky, Amos, and Daniel Kahneman (1974), “Judgment under Uncertainty: Heuristics and Biases”, *Science*, 6(185), pp. 1124–31.

10. Borel, Emile (1909), “Les probabilités dénombrables et leurs applications arithmétiques”, *Rendiconti Circolo Matematico di Palermo*, 27, pp. 247–70. For details on Borel’s work, see Bru, Marie-France, and Bernard Bru, *Les jeux de l’infini et du hasard*, vol. 1., pp. 81–9.

11. Plato, Jan, von, (1994), *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*, (Cambridge Studies in Probability, Induction and Decision Theory), Cambridge, UK: Cambridge University Press.

12. Barbut, Marc, Bernard Locker and Laurent Mazliak (2014), *Paul Lévy and Maurice Fréchet: 50 Years of Correspondence in 107 Letters*, London: Springer-Verlag.

13. Mazliak, Laurent (2015), "The Ghosts of the Ecole Normale. Life, Death and Legacy of René Gateaux", *Statistical Science*, 30(3), pp. 391–412.

14. Cont, Rama, and Peter Tankov (2004), *Financial Modelling With Jump Processes*, Boca Raton, London, New York, Washington, D.C.: Chapman & Hall/CRC., p. 103.

15. Lévy, Paul (1948), *Processus stochastique et mouvement brownien*, Paris: Gauthier-Villars. The text by Lévy is followed by a 50-page note by Michel Loève on the general theory of stationary and related processes.

16. Meyer, Paul-André (2009) [2000], "Stochastic Processes from 1950 to the Present" ["Les processus stochastiques de 1950 à nos jours" in Pier, Jean-Paul (ed.), *Development of Mathematics 1950-2000*, Bale: Birkhäuser, pp. 813–48], translated from the French by Jeanine Sedjro, *Electronic Journ@l for History of Probability and Statistics*, 5(1), June, p. 3.

17. O'Connor, John Joseph, and Edmund F. Robertson (2001), "Kiyosi Ito", MacTutor History of Mathematics Archive, School of Mathematics and Statistics, University of St. Andrews, Scotland:

http://www-history.mcs.st-andrews.ac.uk/Biographies/Ito.html

18. Duplantier, Bertrand, and Vincent Rivasseau (eds) (2014), "Poincaré, 1912–2012", *Proceedings of the Poincaré Seminar*, XVI, 24 November 2012. Progress in mathematical physics, Basel: Birkhäuser.

19. Brown, Robert (1828), "A Brief Account of Microscopical Observations made in the Months of June, July and August, 1827, on the Particles contained in the Pollen of Plants; and on the general Existence of active Molecules in Organic and Inorganic Bodies", *The Philosophical Magazine and Annals of Philosophy*, 4(21), September: pp. 161–173.

20. Bachelier, Louis (2006 [1900]), *Louis Bachelier's Theory of Speculation: The Origins of Modern Finance* [*Théorie de la speculation*, Paris: Gauthier-Villars], Translated and with Commentary by Mark Davis and Alison Etheridge; Foreword by Paul A. Samuelson, Princeton, NJ: Princeton

University Press.

21. Einstein, Albert (1905 [2006]) [Manuscript received: 11 May 1905], written at Berne, Switzerland, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen” [“On the Motion – Required by the Molecular Kinetic Theory of Heat – of Small Particles Suspended in a Stationary Liquid”], *Annalen der Physik*, Berlin, Germany (1905); Hoboken, NJ, USA, 10 March 2006, 322(8), pp. 549–560.

22. Langevin, Paul (1908 [1997]), “Sur la théorie du mouvement brownien” [“On the Theory of Brownian Motion”], *Compte rendu de Académie des Sciences*, Paris, 146, pp. 530–33 [Introduced by Don Lemons, translated from the French by Anthony Gythiel, American Association of Physics Teachers].

23. Smoluchowski, Marian (1916), “Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen”, *Physik Zeitschrift*, vol. 17, pp. 557–571.

24. Lévy, Paul (1922), *Leçons d’analyse fonctionnelle professées au Collège de France*, Paris: Gauthier-Villars.

25. Itô, Kiyosi and Makiko Nisio (1968), “On the oscillation functions of Gaussian processes”, *Mathematica Scandinavica*, vol. 22, pp. 209 –223.

26. Daniell, Percy J. (1918), “A General Form of Integral”, *Annals of Mathematics*, Second Series, 19(4), pp. 279–294.

27. Itô, Kiyosi (1998), “My Sixty Years in Studies of Probability Theory: Acceptance speech of the Kyoto Prize in Basic Sciences”.

## The *Prisme* Series

The *Prisme* Series is a collection of original texts that focus on contemporary theoretical issues. The authors are contributors to the Cournot series of conferences, panels and seminars.

## Latest releases:

***39. Space Surveillance and Risk Assessment: The Economics of Debris***
George Papanicolaou

***38. Fractional and Statistical Randomness***
Charles Tapiero and Pierre Vallois

***36. The Reconciliation of Betting and Measure***
Glenn Shafer

***35. Big Data: A Game Changer for Quantitative Finance?***
Mathieu Rosenbaum

***34. Saturation and Growth: When Demand for Minerals Peaks***
Raimund Bleischwitz and Victor Nechifor

A complete list of publications can be found at
**www.centre-cournot.org**

**Daniel Goroff** is Vice President and Program Director at the Alfred P. Sloan Foundation in New York. He is currently on temporary and part-time loan to the National Science Foundation (NSF) serving as Division Director for Social and Economic Sciences. He is also Professor Emeritus of Mathematics and Economics at Claremont's Harvey Mudd College, where he served as Vice President for Academic Affairs and Dean of the Faculty. He is a Trustee of Smith College and a member of the Board of Directors of the Global Legal Entity Identifier Foundation in Basel. He is also a Distinguished Visiting Professor of Mathematics at Columbia University's Teachers College.

In pursuing his research on nonlinear systems, chaos, and decision theory, Daniel Goroff has held visiting positions at the Institut des Hautes Études Scientifiques in Paris, Mathematical Sciences Research Institute in Berkeley, Bell Laboratories in New Jersey, and Dibner Institute at MIT. He is a Fellow of the American Association for the Advancement of Science in their Mathematics Section.

Series Editor: Jean-Philippe Touffut
Illustration Artist: Bertrand Mac Gaw / Cover Designer: Sophie Otrage