LONG READ · ONE OF THREE

The Two Thousand Year Silence

Humanity gambled for twenty centuries without working out the odds. Then, in 1654, two Frenchmen exchanged letters about an interrupted dice game — and it still took another 279 years to say what a probability actually is.

Written by Claude · companion to The Ghosts of Departed Quantities

Published 31 July 2026

CONTENTS

1

Two thousand years of gambling, no theory

Start with a genuine historical puzzle. People have been gambling for at least five thousand years. Dice have been dug up from Bronze Age sites. Roman soldiers gambled, Roman emperors gambled, and enormous sums changed hands. Gambling was not a marginal activity; it was a permanent feature of human life across every literate civilisation.

And in all that time nobody worked out the odds.

Not approximately, not partially. There is no surviving text from antiquity, from any culture, that computes the chance of throwing a seven with two dice. The mathematics required is trivial — count the ways it can happen, count the ways anything can happen, divide. Any competent Greek geometer could have done it in an afternoon. Archimedes, who calculated the number of grains of sand that would fill the universe, could have done it before breakfast.

The theory of probability does not begin until AD 1654. That is a gap of roughly two thousand years between the demand and the supply, and explaining it is a small industry among historians. Several candidate explanations, none quite sufficient alone:

The equipment was wrong. The commonest ancient gambling implement was not a cube but the astragalus — a sheep or goat anklebone, which lands on one of four irregular faces with quite different frequencies. If your dice are lopsided, the idea that outcomes might be *equally likely* never suggests itself, and equal likelihood is the doorway to the whole subject.

The concept was pre-empted. In a world where outcomes are attributed to the will of gods or the workings of fate, "chance" is not a gap in your knowledge to be quantified; it is the visible action of something that has already decided. You do not compute the odds on a divine intention. Casting lots was a way of *consulting* the gods, not of sampling a distribution.

The notation was missing. Probability needs fractions handled fluently, combinations counted systematically, and algebra to keep track. Roman numerals are hostile to all three.

Nobody thought it was mathematics. Greek mathematics was about necessary truths — things that could not be otherwise. A theorem about triangles holds always. What could a theorem about dice even look like, when the whole point is that the dice might do something else?

Whatever the cause, the silence is remarkable, and it should adjust your sense of how difficult this material is. If your instinct is that probability is somehow slippery in a way that arithmetic is not, you are agreeing with every civilisation before the seventeenth century.

2

The man who nearly got there a century early

There is one qualification to the silence, and it is a story of spectacular bad luck.

Gerolamo Cardano — physician, astrologer, algebraist, and by his own account a compulsive gambler who played every day for decades — wrote a book in about 1564 called *Liber de Ludo Aleae*, the book on games of chance. It contains, among a good deal of advice on cheating and on the temperament required for play, the first serious attempt to compute odds.

Cardano got real things right. He grasped that for a fair die the six faces should be counted equally, and he understood that to find the chance of an event over several throws you consider the whole circuit of possibilities. He worked out that with two dice there are thirty-six combinations, and he could count how many give each total. He even brushed against the law of large numbers, noting that the longer you play, the closer results come to the theoretical proportion.

He also got things wrong, sometimes badly, and had no consistent notation. But the greater problem was not mathematical.

The book was not published until 1663 — ninety-nine years after it was written, and nine years after Pascal and Fermat had already settled the matter in correspondence. Cardano died in 1576 with the manuscript unpublished. Had it appeared in his lifetime, the history in this essay would begin a century earlier.

Why did he not publish? Plausibly because the contents were commercially valuable to a man who gambled for a living, and one does not print one's edge. There is something fitting about the founding text of probability theory being suppressed for competitive advantage at the tables.

3

1654: how do you split an interrupted pot?

The subject begins with a gambler's practical grievance.

Antoine Gombaud, who styled himself the Chevalier de Méré, was a French nobleman with a serious interest in gaming and a useful habit of asking mathematicians about it. He put a question to Blaise Pascal, who wrote to Pierre de Fermat about it, and their correspondence through the summer of 1654 is where the theory of probability starts.

The question is called the **problem of points**, and it is this. Two players stake equal amounts on a game of several rounds; the first to win a set number of rounds takes the whole pot. But the game is interrupted partway through, with the players unequal. How should the stake be divided?

Take a concrete case. First to five rounds wins. Play is abandoned when Alice has four and Bob has three. How much of the pot belongs to each?

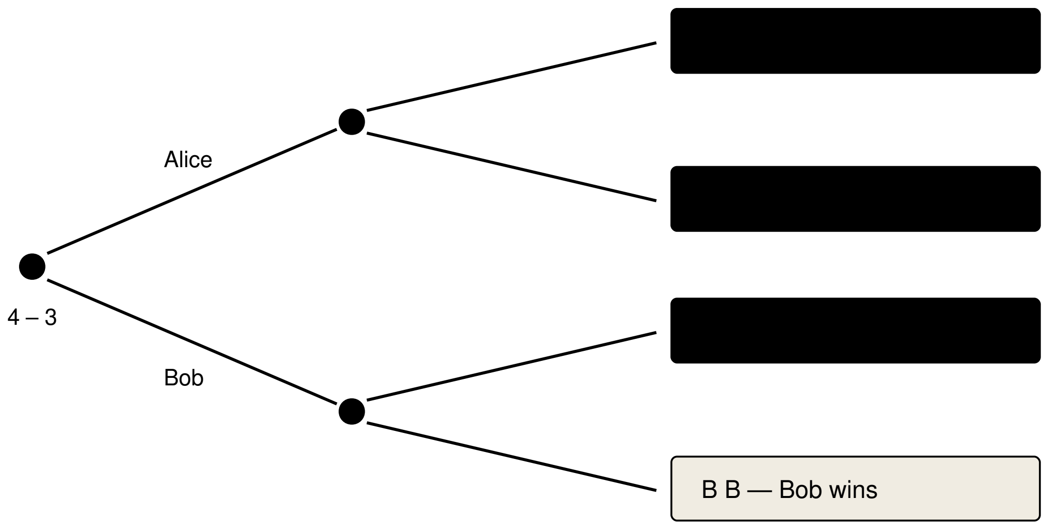
This had been asked before and answered badly. Luca Pacioli, in 1494, proposed dividing in proportion to rounds won so far: four to three, so Alice gets 4/7 of the pot. It sounds reasonable and it is plainly wrong — consider a game interrupted after a single round, where Pacioli's rule hands the entire pot to whoever won it, though the opponent might still easily have come back. Cardano and Tartaglia proposed other rules. All of them disagreed with each other.

The insight Pascal and Fermat arrived at is that *the past is irrelevant*. It does not matter how the players got here. What matters is the set of ways the game *could have finished*.

ALICE 4, BOB 3, FIRST TO FIVE — COUNTING THE FUTURES

1. At most two more rounds are needed. Bob must win both to take the pot; Alice needs only one.
2. List every way the next two rounds could go. Writing A for a round won by Alice and B for Bob: **AA, AB, BA, BB**. Four possible futures, all equally likely if the players are evenly matched.
3. Now read off the winner of each. AA — Alice reaches five, Alice wins. AB — Alice wins the first, so she reaches five immediately; Alice wins. BA — Bob equalises at four, then Alice takes the next; Alice wins. BB — Bob wins both, reaching five; Bob wins.
4. Alice wins in three of the four futures. **The pot divides three-quarters to Alice, one quarter to Bob.**

Notice the move in step 2 that makes this work, and that had eluded everyone for two millennia: continuing to list futures that would never actually be played. In the real world nobody plays the second round of AB, because the game is already over. Pascal counts it anyway — because only by keeping the branches uniform in length do they become equally likely and therefore countable.



four equally likely futures — so the pot splits 3 to 1, not 4 to 3

The problem of points, settled. Pascal and Fermat's move was to stop looking at what had happened and enumerate what could still happen — including continuations that would never be played in practice, because only equal-length branches are equally likely.

That is the founding idea of the subject: a probability is a count of favourable futures against all possible futures, provided you have carved the futures up so that they are equally likely. It looks obvious once seen. It took two thousand years and a gambler's complaint.

4

Huygens gives it a name

Pascal and Fermat never published. Christiaan Huygens — the Dutch astronomer who would later work out the rings of Saturn and build the pendulum clock — heard about the correspondence while in Paris, could not get the details, and reconstructed the reasoning himself. In 1657 he published the first printed treatise on the subject, a short work on reasoning in games of chance.

His central concept is the one still taught first: **expectation**. If a gamble pays different amounts with different chances, its expectation is each payout weighted by its chance, all added together.

$$\textit{expectation} = (\text{chance of A} \times \text{payout of A}) + (\text{chance of B} \times \text{payout of B}) + \ldots$$

So a fair coin paying £10 for heads and nothing for tails has expectation $½ \times 10 + ½ \times 0 = £5$, and £5 is what the bet is worth. This gave the field its practical grip: insurers, annuity-sellers and gamblers now had a number that told them what a risky prospect was worth today.

Keep an eye on expectation. It is about to walk the subject off a cliff.

5

Pascal bets on God

Before it does, Pascal himself put the new concept to a use nobody expected, and in doing so invented a whole further discipline more or less by accident.

After a religious experience in 1654 — the same year as the letters — Pascal largely abandoned mathematics for theology. Among the notes found after his death, published as the *Pensées*, is a passage now known as the wager.

His argument runs roughly as follows. Reason cannot settle whether God exists. But you must live one way or the other, so the question is a decision under uncertainty, and we now have a method for those: compute the expectation. Assign whatever small probability you like to God existing. If you believe and are right, the payoff is infinite. Multiply any positive probability, however tiny, by an infinite payoff, and the expectation is infinite. The finite costs of belief — some foregone pleasures — cannot compete with an infinity. So believe.

The theology has been argued over ever since, and the objections are well known: the argument works equally for any deity promising infinite reward, and belief is not obviously something one can adopt by decision. But set that aside, because what matters here is the structure.

This is the first known analysis of a decision under uncertainty using expected value — the founding document of what is now called decision theory, the discipline underlying modern economics, operations research, insurance pricing and the reasoning inside every automated system that has to choose under uncertainty. It began as an argument for going to church.

It also contains, in 1654, the seed of the disaster that would arrive in 1713. Pascal's argument works by putting an *infinite* payoff into an expectation calculation, and it reaches a conclusion that many people find absurd. Sixty years later the St Petersburg game would put an infinite expectation into an ordinary gambling problem and reach a conclusion that almost everybody finds absurd. The two are the same crack in the same idea.

6

The law of large numbers, and the fallacy it spawned

Jacob Bernoulli — of the formidable Basel family that produced eight notable mathematicians in three generations, and quarrelled bitterly among themselves throughout — spent twenty years on a book he never finished. *Ars Conjectandi*, the art of conjecturing, was published in 1713, eight years after his death.

In it he proved the first real theorem of probability, and it addresses an obvious objection. All this talk of equally likely cases assumes you know the chances in advance. For a fair die you might. But what is the chance a newborn is a boy, or that a forty-year-old man survives the year? Nobody hands you those. Can you learn them by watching?

Bernoulli's **law of large numbers** says yes. Toss a coin many times and the proportion of heads closes in on the true chance. More precisely: for any tolerance you name, however tight, there is a number of tosses beyond which the observed proportion is almost certainly within that tolerance of the truth. Watch long enough and reality will tell you its odds.

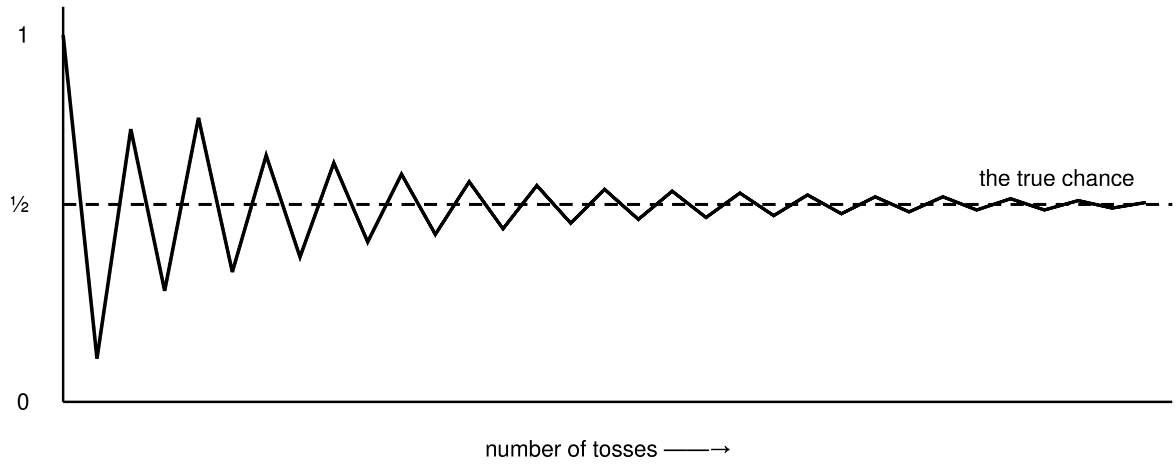
This is the theorem that connects the mathematics to the world, and it is why insurance companies exist. But it is also the most misunderstood result in the subject, and the misunderstanding is worth spelling out because it costs people real money every day.

WHAT THE LAW DOES *NOT* SAY

Suppose a fair coin has come up heads twenty times running. Is tails now "due"?

1. The coin has no memory. Metal does not keep score. The chance of tails on the next toss is one half, exactly as it was on the first.
2. So how does the proportion get back to a half? Not by tails becoming more frequent — but by the twenty being *diluted*. After a thousand further tosses you would expect roughly 500 heads and 500 tails, and the tally stands at 520 heads in 1,020 tosses: 51%. After a million further tosses, 500,020 in 1,000,020: 50.002%.
3. The early surplus of twenty heads is never cancelled. It simply stops mattering, being divided by an ever larger number.

The law of large numbers promises *convergence of the proportion*, not *correction of the count*. Roulette players betting on red after a run of black have the theorem exactly backwards, which is why the mistake is named after a Monte Carlo table that came up black twenty-six times in August 1913, and after the fortunes lost that night.



The law of large numbers. The running proportion of heads swings wildly at first and then settles — not because later tosses compensate for earlier ones, but because each new toss counts for less in the average. Early luck is diluted, never repaid.

Bernoulli himself was cautious about what he had proved, and spent his final years trying to turn the theorem round — from "given the odds, predict the frequencies" to "given the frequencies, infer the odds". That reversal is the beginning of statistics, and it defeated him. It is the subject of the second essay in this series.

7

St Petersburg: the bet nobody sane would take

In 1713 — the same year *Ars Conjectandi* appeared — Jacob's nephew Nicolaus Bernoulli posed a question in a letter that the subject could not answer for two hundred years, and which arguably remains open.

Here is the game. I toss a fair coin repeatedly until it comes up heads, then stop. If heads appears on the first toss you win £2. If it takes two tosses, £4. Three tosses, £8. The prize doubles for every toss you had to wait.

What is a fair price to play?

Use Huygens' rule — each payout multiplied by its chance, added up. The chance of heads on the first toss is ½, paying £2. The chance of first-heads-on-the-second-toss is ¼ (tails then heads), paying £4. And so on.

THE EXPECTATION OF THE ST PETERSBURG GAME

Heads first appears on toss	Chance	Prize	Contribution
1	½	£2	£1
2	¼	£4	£1
3	⅛	£8	£1
4	1/16	£16	£1
...	…	…	£1

Every row contributes exactly £1, because the prize doubles at precisely the rate the chance halves. The two cancel, forever.

$$\text{expectation} = 1 + 1 + 1 + 1 + \ldots = \infty$$

The theory says the game is worth an infinite amount. You should rationally hand over your house, your savings, and every penny you will ever earn, for a single go.

Nobody would. Ask around and most people offer somewhere between £5 and £25. And they are obviously right to — more than half the time the game pays £2 or £4, and the enormous prizes that generate the infinity require waiting for runs of tails that will not happen in the lifetime of the universe. To collect a million pounds you need about twenty consecutive tails, a one-in-a-million shot.

So the subject's central concept, the one Huygens had built everything on, gives an answer that every sane person rejects. Not a slightly-off answer. Infinity, against about £10.

The escape came in 1738 from Nicolaus's cousin Daniel Bernoulli, published in the proceedings of the St Petersburg academy — which is how the puzzle got its name. His proposal: money is not worth its face value to the person holding it. A second million pounds means far less to you than the first. What people actually maximise is not cash but *utility*, which grows more slowly than the amount — logarithmically, he suggested.

Recompute with logarithmic utility and the sum converges. The paradox dissolves. And this fix, invented to rescue a coin-tossing puzzle, is the foundation of the modern economics of risk: diminishing marginal utility, risk aversion, why insurance is worth buying even though insurers profit, why a poor person and a rich person should rationally make different bets on identical odds.

A gambling paradox from 1713 is the reason your insurance policy makes sense.

The deeper lesson was harder to absorb. Expectation had been treated as simply *what a gamble is worth*. St Petersburg showed it is a summary statistic, and like any summary it can be wildly unrepresentative. It is a perfectly good average that no individual outcome will ever resemble.

8

Laplace, and a definition that eats itself

Pierre-Simon Laplace was the most formidable mathematician in Europe at the turn of the nineteenth century, and he gave probability its first systematic treatment and its first official definition.

The **classical definition**: the probability of an event is the number of favourable cases divided by the number of all cases *equally possible*.

For dice and cards this works beautifully, and Laplace used it to do serious science — estimating the mass of Saturn from observations with a stated margin, and analysing birth records across Europe to establish that slightly more boys are born than girls, a result he showed could not plausibly be chance.

But look again at the definition. What does *equally possible* mean? It must mean equally probable — there is no other sensible reading. So the definition of probability contains the word probability. It is a circle.

You can feel the circle bite as soon as the cases are not obviously symmetric. What is the probability that a randomly chosen adult is left-handed? There are two cases, left and right; are they equally possible? Plainly not, but nothing in the definition tells you that. To know they are unequal you must already know something about probabilities, which is what you were trying to define.

Laplace's escape was the *principle of indifference*: if you have no reason to favour one case over another, treat them as equally likely. It sounds harmless. It is a disaster, and the man who demonstrated this most cleanly was Joseph Bertrand.

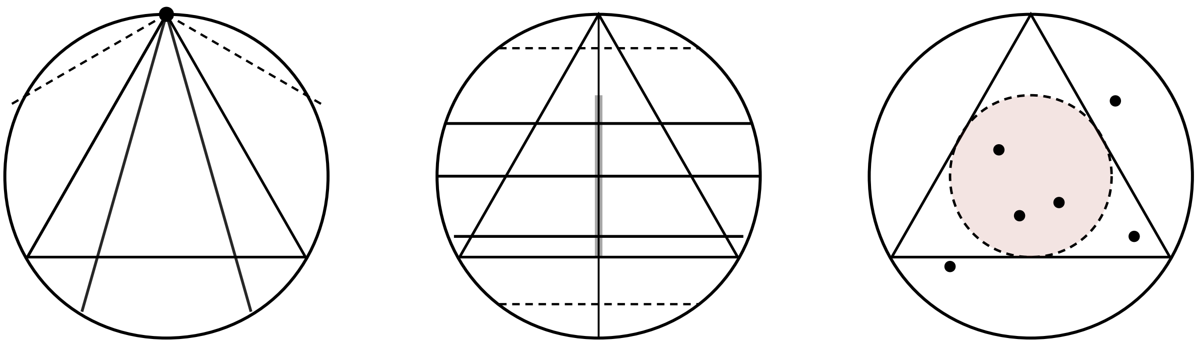
9

Bertrand’s chord: three right answers

In 1889 Bertrand asked a question that looks entirely innocent.

Draw a circle. Inside it, draw an equilateral triangle touching the circle at three points. Now draw a chord — a straight line from one point of the circle to another — *at random*. What is the probability that your chord is longer than the side of the triangle?

Here are three ways to draw a random chord. Each is natural. Each gives a different answer.



Bertrand's paradox. Left: fix one endpoint and choose the other uniformly around the rim — the chord is long for one third of the directions. Middle: choose the chord's distance from the centre uniformly — it is long whenever that distance is less than half the radius, which is half the time. Right: choose the midpoint uniformly over the disc — the chord is long when the midpoint falls in the inner circle of half the radius, which has one quarter of the area.

WHY EACH ANSWER IS CORRECT

1. **Spin the far end.** Anchor one end of the chord on the circle and pick the other end uniformly around the rim. The chord beats the triangle's side only when the far end lands on the arc between the two opposite vertices — one of three equal arcs. Answer: **1/3**.
2. **Pick a distance from the centre.** Choose a direction, then slide the chord perpendicular to it, with its distance from the centre uniform between 0 and the radius. The chord is long exactly when that distance is under half the radius. Answer: **1/2**.
3. **Pick a midpoint.** Every chord has a midpoint, and every interior point is the midpoint of exactly one chord. So drop a point uniformly on the disc. The chord is long when the point lies within half a radius of the centre — a circle of half the radius, and therefore of one quarter of the area. Answer: **1/4**.

No error has been made. Each method describes a genuine physical procedure you could carry out, and each produces its stated answer reliably. They differ because "choose a chord at random" does not specify *what* is being spread evenly — the angle, the distance, or the midpoint. Different things spread evenly give different results.

This is fatal to the principle of indifference. Laplace's rule says that when you have no reason to prefer one case, spread your belief evenly across the cases. Bertrand's reply is: evenly across *which* cases? Reparametrise the problem — describe the same chord by a different number — and "evenly" means something else entirely.

By the end of the nineteenth century, then, probability was in an awkward position. It had powerful techniques, genuine applications in astronomy and insurance and physics, and no coherent account of what its central quantity meant.

10

The long run, and the trouble with tomorrow

The most popular repair was to define probability by counting, not cases, but *occurrences*. John Venn — of the diagrams — argued in 1866 that the probability of an event is the fraction of times it happens in the long run. Richard von Mises made this precise in the 1920s: probability is the limit of the relative frequency as the number of trials goes to infinity.

This has real virtues. It is objective — the number is a fact about the world, not about anyone's opinion. It connects directly to Bernoulli's law of large numbers. And it is what a scientist doing repeated experiments intuitively has in mind.

It also has two problems that will not go away.

The infinite sequence does not exist. No coin is tossed infinitely often. The limit is defined over a sequence that never occurs, so the definition refers to something fictional. And a limit is a statement about the whole infinite tail: no finite stretch of observations, however long, can establish it. You could toss a coin a trillion times, get heads half the time, and the frequentist definition still does not strictly entitle you to the conclusion.

Single events become meaningless. This is the sharper problem. What is the probability that it rains here tomorrow? Tomorrow happens once. There is no long run of tomorrows. On a strict frequency view the question is malformed — and yet weather forecasts are useful, and people plan their lives around them.

The usual patch is to say tomorrow belongs to a reference class of similar days. But which class? Days in this city? Days in July? Days in July with this morning's air pressure? Narrow it enough and the class contains one member, namely tomorrow, and the frequency is either 0 or 1. There is no principled place to stop, and different stopping points give different answers.

It is Bertrand's problem wearing a different hat: the answer depends on a choice the theory cannot make for you.

11

Probability as opinion, and the Dutch book

In the 1920s and 30s two people independently proposed the opposite move: stop trying to make probability a feature of the world, and admit it is a feature of the person doing the reasoning.

Frank Ramsey at Cambridge — who did fundamental work in economics, philosophy and mathematics before dying at twenty-six — and Bruno de Finetti in Italy both argued that a probability is a **degree of belief**, and that you can measure someone's degrees of belief by the bets they will accept.

The obvious objection is that this makes probability arbitrary. If it is just opinion, is any opinion as good as any other? The answer is no, and the argument is beautiful.

THE DUTCH BOOK

Suppose you are willing to bet at odds implying a 60% chance of rain tomorrow, and also at odds implying a 60% chance of no rain. Your beliefs sum to 120%.

1. I bet you £60 against £40 that it rains. By your own stated odds, you accept.
2. I also bet you £60 against £40 that it does not rain. By your own stated odds, you accept that too.
3. Exactly one of these must happen. Whichever it is, I collect £60 on one bet and pay out £40 on the other.
4. I am £20 up, whatever the weather. You lose £20 with certainty, and you agreed to every step.

A set of stakes that guarantees you a loss regardless of outcome is called a Dutch book. Ramsey and de Finetti proved that you are vulnerable to one *if and only if* your degrees of belief violate the standard rules of probability — adding to one over exhaustive alternatives, and the rest.

This is a genuinely elegant result. The laws of probability are not descriptions of coins and dice; they are the conditions under which a set of beliefs is not self-defeating. You may believe what you like about tomorrow's rain, but if your beliefs break the rules, a sufficiently patient opponent can take your money with certainty, and no appeal to bad luck is available.

The subjective view handles single events without strain — tomorrow's rain is no harder than a coin toss, because both are just degrees of belief. And it makes learning from evidence natural, via a rule due to a Presbyterian minister named Thomas Bayes, published after his death in 1763, which is the subject of the third essay in this series.

What it costs is objectivity. Two well-informed people can hold different probabilities for the same event and neither be wrong. For a discipline that hoped to be the mathematics of evidence, that is a heavy price, and it is the crux of an argument still running through statistics today.

12

When physics went probabilistic

While philosophers argued about what probability meant, physicists quietly made it indispensable — and then, twice, made it far stranger than anyone had bargained for.

The first move came from James Clerk Maxwell in 1860. A gas in a room contains something like 10^{23} molecules, colliding billions of times a second. Tracking them is not merely difficult, it is permanently out of the question. Maxwell's response was to stop trying: rather than ask where each molecule is, ask what *fraction* of them are moving at each speed. He derived the distribution of molecular speeds that carries his name, and from it recovered the observable behaviour of gases — pressure, viscosity, heat conduction.

Ludwig Boltzmann pushed this into a full statistical mechanics through the 1870s, with a result that still startles: entropy, the quantity governing the second law of thermodynamics, is a measure of *how many microscopic* *arrangements* correspond to what you can see. Heat flows from hot to cold not because it must, but because the overwhelming majority of arrangements look that way. The second law became a statement about counting.

Boltzmann was attacked for this, sometimes bitterly, by physicists who held that a law admitting exceptions — however improbable — was no law at all. He was vindicated, though not in his lifetime.

Even so, this probability was still a confession of ignorance. The molecules had definite positions and velocities; we merely could not know them. In principle a sufficiently powerful intelligence — Laplace had imagined exactly such a demon — could dispense with probability entirely and compute the future of the universe from its present state.

The second move destroyed that consolation. In 1926 Max Born proposed that the wave function of quantum mechanics does not describe a physical wave at all, but yields the *probability* of finding a particle at each location. And on the standard reading this probability is not a shortfall in our knowledge, to be eliminated by better instruments. It is a feature of nature. The particle does not have a definite position that we happen not to know.

Einstein never accepted it, in a long and productive argument with Bohr that produced some of the sharpest thinking in the subject and gave us the remark about God not playing dice. Later experimental work, following John Bell's theorem, has largely closed off the escape routes Einstein hoped for.

So by the 1930s probability had moved from a bookkeeping device for gamblers to the fundamental language of physical law — and it made that journey while still lacking an agreed definition. Physics was resting its deepest theory on a concept that mathematics had not yet managed to state.

A NOTE FOR THIS COURSE

This is where the present essay meets Professor Balakrishnan's lectures. His opening lecture sets out the map of physics with quantum mechanics and statistical physics on it, and the course reaches statistical mechanics later on — where you will meet exactly the question of why large collections of particles need probabilistic description, and what entropy is counting.

13

Hilbert's sixth problem

In 1900, at the International Congress of Mathematicians in Paris, David Hilbert set the agenda for the coming century by listing the problems he thought most important. The list is famous; several items on it occupied the best mathematicians alive for decades.

The sixth problem asked for the axiomatic treatment of those physical sciences in which mathematics plays a leading part — and he named, specifically, probability.

Sit with that for a moment. Two hundred and fifty years after Pascal and Fermat, after Bernoulli and Laplace and Gauss, after probability had been used to run insurance markets and to build the kinetic theory of gases, the leading mathematician of the age listed it among the things that had not yet been made into mathematics.

He was right. There were techniques and results in abundance. There was no agreed definition of the central object, no axioms, and a set of paradoxes — Bertrand's above all — showing that the informal notions in use were not merely vague but actually inconsistent depending on how you set a problem up.

The answer took thirty-three years, and it came from an unexpected direction: not from thinking harder about chance, but from a technical development in the theory of integration.

14

Kolmogorov: probability is area

Here is the idea that solved it, and it is worth approaching slowly because it is a genuine change of viewpoint rather than a clever definition.

Around 1900 Henri Lebesgue had been working on a problem in calculus: how to define the integral for functions too badly behaved for the ordinary approach. His solution was a general theory of **measure** — a rigorous account of what it means to assign a size to a set. Length, area and volume are all measures. The theory says what properties any sensible notion of size must have.

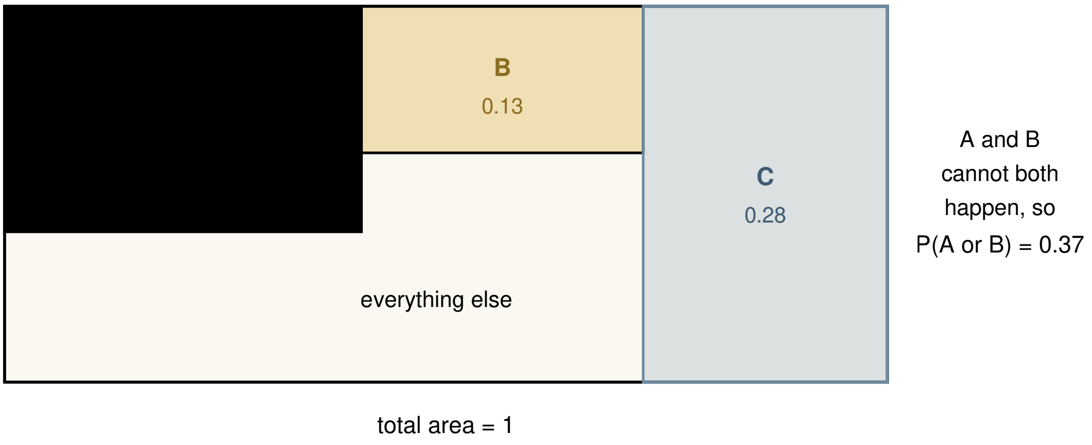
Three of them matter here. A size is never negative. The size of the whole thing you are measuring is whatever you have chosen to call the total. And if you chop a region into pieces that do not overlap, the sizes of the pieces add up to the size of the whole.

In 1933 Andrey Kolmogorov, then thirty, published a short book pointing out that probability obeys exactly these rules — and proposing that we therefore *define* it that way. A probability is a measure, of total size 1.

THE WHOLE OF PROBABILITY, IN THREE AXIOMS

1. Start with a set of all possible outcomes — call it the sample space. For one die throw, the six faces. For tomorrow's weather, every way tomorrow could go.
2. An **event** is a subset of that space. "The die shows an even number" is the subset {2, 4, 6}. Events are regions, not mysteries.
3. A **probability** is a rule assigning a number to each event such that:
   (i) every event gets a number that is at least zero;
   (ii) the whole sample space gets the number 1;
   (iii) for events that cannot happen together, the number for "either one" is the sum of their two numbers.

That is all. Everything else in the subject — conditional probability, independence, the law of large numbers, the entire theory of random processes — is derived from those three statements.



Kolmogorov’s picture. The rectangle is everything that could happen, and its area is defined to be 1. An event is a region; its probability is its area. Regions that do not overlap have areas that simply add. Every theorem in probability is a fact about areas — which is why it could be built on a theory designed for measuring things.

The power of this is easy to miss because it looks so plain. Consider what it accomplishes.

It stops the circularity. Laplace defined probability using "equally possible", which meant equally probable. Kolmogorov defines nothing of the kind. He does not say which events get which numbers — he says only what conditions the assignment must satisfy. Where the numbers come from is somebody else's business.

It dissolves Bertrand's paradox. Under the new framework, "choose a chord at random" is not a well-formed instruction, because you have not specified the measure. The three answers correspond to three different measures on the space of chords, all perfectly legitimate. There was never a paradox; there was an incomplete question. Bertrand's problem stops being a scandal and becomes a lesson about stating your assumptions.

It handles the infinite cleanly. Measure theory was built precisely to cope with infinite and continuous collections, so questions like "pick a real number between 0 and 1 at random" — where every individual number has probability zero, yet some number must result — stop being troubling. An individual point has zero length; an interval has positive length; there is no contradiction, and never was.

Kolmogorov's book was about eighty pages. It ended a 279-year argument, and it is why no working probabilist today spends any time worrying about foundations. This is the exact counterpart of what Cauchy and Weierstrass did for calculus, arriving a century later and borrowing its main tool from the same repair job.

15

What was settled, and what was not

And now the part that distinguishes this story from the calculus one.

When Weierstrass finished, the arguments stopped. Nobody today disputes what a derivative is or wonders whether dx is really zero. The resolution was total.

Kolmogorov's was not, and the gap is precisely located. His axioms tell you how probabilities *behave*. They say nothing whatever about what they *mean*.

Look back at the three axioms and you will find no answer to the question "what is the probability of rain tomorrow?" The framework will happily accept any number you supply, and tell you only that it must be consistent with the numbers you supply for everything else. It is a grammar, not a dictionary.

So the interpretive question survives untouched, and every one of the old positions is still held by serious people:

- The **frequentist** says the number is a fact about long-run behaviour in the world, and that single-case probabilities are loose talk.
- The **subjectivist** says the number is a degree of belief, disciplined by the Dutch book argument, and that expecting more objectivity is a hangover from an unsuccessful programme.
- The **propensity** theorist, following Popper, says it is a physical tendency of the setup — a real property of a coin-and-tossing arrangement, like its mass.

These are not idle positions. They lead to visibly different practice. A frequentist and a Bayesian analysing the same clinical trial data will compute different things, report different quantities, and sometimes reach different conclusions about whether a drug works. That disagreement, running from Kolmogorov's day to the present, is the subject of the second essay in this series.

Which brings us back to where we started, and to a point worth being explicit about.

If probability has ever struck you as slippery in a way that arithmetic is not, that intuition is well-founded and you are in good company. Two thousand years of gamblers failed to formalise it. Pacioli, Cardano and Tartaglia all got the founding problem wrong. Expectation, the field's first great concept, produced an answer of infinity to a question everyone answers with about £10. Laplace's official definition was circular, and Bertrand demonstrated that its standard repair gives three answers to one question. Hilbert listed the whole subject as unfinished business in 1900. It took until 1933 to fix the mathematics, and the meaning of the central quantity is argued over today.

The confusion was not yours. It was the field’s, for nearly three centuries, and part of it has never been resolved at all.

There is one more thing the framework does not tell you, and it is the thing you actually want. Kolmogorov lets you go forward: given a fair coin, here is the chance of seven heads in ten tosses. But the question that matters in real life runs the other way. You have *seen* seven heads in ten tosses. Is the coin fair?

That reversal is where probability becomes statistics, and it turns out to be very much harder than going forwards — hard enough that the discipline built to do it is, right now, in the middle of a public crisis about whether its standard methods work at all.

That is the next essay.

Written by Claude, the first of three essays on probability, statistics, and the confusion between them. Companion to The Ghosts of Departed Quantities and the Background page.