

THE IN A HURRY SERIES

Probability

in a Hurry

Odds, risk, and the maths of chance

Max Brocklesby

BOOKSINAHURRY.COM

Contents

The Whole Thing in One Page 4

Why You Should Care 5

The Core Ideas 7

How It Actually Works 21

What People Get Wrong 30

Use It 35

Terms 40

Go Deeper 45

Notes and Sources 47

Bibliography 55

The Whole Thing in One Page

Probability is often introduced as arithmetic for dice and then treated as a dim view of the future. Both pictures miss the central act. A probability attaches weight to an event inside a model, given what is known. It may describe physical chance, a long-run pattern, evidential support or a coherent degree of belief. The arithmetic is shared even when interpretations differ.

The model comes first. A fair die has six face outcomes because face, rather than angle, bounce or sound, is what the question records. Symmetry supports equal weights. The model does not create the object's physical tendencies, but it decides which results count and how evidence or mechanism connects to them. Change the event, information or horizon and the probability may change. A clean-looking 70 per cent carries an unwritten clause.

Counting is the first engine. If elementary outcomes are equally likely, probability becomes favourable cases divided by possible cases. That is how cards, combinations and the birthday problem work. Twenty-three people are enough to make at least one shared birthday more likely than not, under the standard simplified model, because the question compares every pair with every other pair. Rare per pair does not mean rare across many opportunities.

Information is the second engine. Conditional probability changes the denominator from all possibilities to those still compatible with what you have learned. That is why independence is a claim rather than a feeling, and why the Monty Hall switch wins two times in three only when the host follows the stated rule. Bayes' theorem turns the same machinery around. It combines what was plausible before with how strongly new evidence favours one explanation over another. A striking test result can remain weak evidence when the target is scarce and false alarms have many chances to appear.

Distributions are the third engine. An uncertain quantity has a range of

possible values, not one forecast. Expected value locates its weighted centre, but two risks with the same mean can have different spread, tails and chances of ruin. Repetition may make an average settle through the law of large numbers. It does not make the next toss compensate for the last one, guarantee a smooth path or rescue someone who has already lost the ability to continue.

Risk begins when probabilities meet consequences. A one-in-a-million failure faced once is remote. Faced independently a million times, its chance of appearing at least once is about 63 per cent. Dependence can make several protections fail together. A favourable long-run average can still be intolerable when losses are irreversible, survival is required or one bad branch closes every future branch.

The history follows the same movement. Gamblers and mathematicians learned to count unfinished games. Bernoulli connected theoretical chance to repeated frequency. Bayes and Laplace formalised learning from evidence. Kolmogorov supplied axioms broad enough for modern mathematics. Computers then made it possible to explore complicated models by simulation when algebra could not finish the job.

Probability therefore offers disciplined uncertainty, not certainty by decimal point. Its power comes from forcing the event, model and information into view. Its danger comes from making a thin model look complete. Good probabilistic thinking asks four questions: what can happen, how is weight assigned, what information changes the weights, and what do the outcomes cost?

That is the book.

Why You Should Care

Suppose a weather forecast gives rain a 70 per cent probability and the day stays dry. Was the forecast wrong?

One dry day does not establish that a 70 per cent forecast was poor, though it can still carry evidence. Forecast quality is assessed across defined cases

using calibration and scores. Among cases assigned about 70 per cent, rain should occur about seven times in ten if the system is calibrated. One outcome can favour one forecast over another, but cannot reveal a calibration rate. The day is binary. The uncertainty beforehand was not.

That distinction reaches beyond weather. Surgeons, engineers, insurers, investors, jurors and project managers face decisions whose outcomes later look obvious. Before the result, several paths are open. After it, one path survives and hindsight erases the rest. Probability keeps the discarded paths visible long enough to make a decision without pretending to know which one will become history.

It also stops large numbers from bullying you. “A 99 per cent accurate test” sounds decisive until accuracy is separated into the chance of detecting a real case and the chance of a false alarm. “One chance in a million” sounds negligible until there are millions of opportunities. “An average return of 5 per cent” sounds like an experience, although the route may contain losses large enough to end the plan. Percentages become useful only after you know the reference class, time period, dependence and consequence.

Probability gives you several compact tools for doing that work. Odds express probability as relative weight and, under conventions, as quoted prices. Counting exposes how opportunities multiply. Conditional probability forces you to say what information is already known. Bayes shows how base rates and evidence interact. Distributions keep the tails visible. Expected value combines possibilities without pretending they are equally painful. The law of large numbers explains why stable averages can emerge from unstable events, while also showing why an individual may never experience the average promised by a long run.

The subject changes how you read certainty. A precise number can come from a weak model. A wide range can be the honest product of strong work. Two people can agree on every calculation and disagree rationally because they assign different starting weights, face different consequences or cannot survive the same loss. Mathematics can make the disagreement clearer without making the decision for them.

Probability also marks the boundary with statistics. Probability usually starts with a model and asks what outcomes it produces. Statistics starts with outcomes and asks what model, parameter or explanation they support. The traffic runs both ways. A probability model tells you what evidence would look surprising. Data then test, refine or reject the model. Neither direction is safe when assumptions disappear from view.

It disciplines counterfactual thinking too. When a project succeeds, the result does not prove the original risk was small. When a decision fails, the loss does not prove the decision was poor. A sound choice can meet a bad branch, while a reckless choice can be rescued by luck. Judging the decision means returning to what was knowable before the outcome selected itself.

The deepest reason to care is that uncertainty is often more than a temporary defect in knowledge. Some reflects ignorance and may shrink with better evidence. Some arises from physical processes for which chance is part of the best available account. Some comes from systems too complicated to track in full, and some from deliberate abstraction because a useful model cannot contain everything. In every case, action arrives before complete knowledge.

You therefore need more than confidence. You need a way to distribute belief across possibilities, revise it without panic, and choose while remembering that low probability is not zero and high probability is not fate. Once that habit is learned, odds stop being decoration and risk stops being a mood. They become questions that can be inspected.

The Core Ideas

Probability Begins with a Model

Before calculation, specify the event and the model that carries it.

Roll a die onto a table. The ordinary model records one of six upper faces, ignoring where it lands, how it bounces and what sound it makes. That is

disciplined omission. A useful model discards detail without discarding what drives the answer. The sample space contains the outcomes treated as distinct; an event collects any of them. “An even number” contains 2, 4 and 6.

Weights then have to be assigned. For an ideal fair die, symmetry supports one-sixth for each face. The basic rules of probability are spare. No probability is negative. The whole sample space has probability one. For a countable collection of events that cannot occur together, the probability that one of them occurs is the sum of their probabilities. From those rules follows the familiar fact that an event and its complement add to one.

The rules constrain weights but do not determine their source or meaning. They may describe objective physical chances, stable frequencies, evidential support or coherent degrees of belief. These interpretations answer different questions through the same calculus. Even objective chance needs a model connecting mechanism to a defined event.

The separation has a cost. A neat answer may depend more on how the experiment was described than on the object itself. Consider a random chord drawn across a circle. Ask for the probability that it is longer than a side of an inscribed equilateral triangle. Different reasonable methods of choosing a chord give different answers. Choose two random points on the circumference and the answer is one-third. Choose a random midpoint uniformly across the circle and it is one-quarter. Choose a random distance from the centre along a fixed radius and it is one-half. The phrase “random chord” did not specify one experiment, so it did not specify one probability.

That is the modelling lesson. Probability is not a property of the word “failure” or “rain” alone. It belongs to a defined event within a model and information state. Before arithmetic, ask: Is the sample space appropriate? Are cases equally weighted? Do the weights come from symmetry, mechanism, frequency, evidence or judgement? Does the model include every material route to the event?

Odds are another language for the same weights. A probability of $3/4$ means

odds of three to one in favour: three units of favourable weight for every one unit against. Odds of a to b convert to probability a divided by a plus b . Decimal betting odds and commercial prices add further conventions, but the mathematical odds concern relative weight, not profit.

Continuous models add one final surprise. Choose a point uniformly from a line segment. The probability of selecting any exact point is zero, yet some point must be selected. Zero here means no single point receives a positive share of a continuous total. It does not mean logical impossibility. Events such as landing within an interval gain probability through length, area or another measure.

The first discipline of chance is therefore not calculation. It is specification. State the event, define the model, explain the weights, and remember that the answer is conditional on those choices and the information available.

Counting Turns Possibilities into Odds

When elementary outcomes are equally likely, probability becomes a counting problem. The formula looks childish: favourable cases divided by possible cases. The difficulty lies in counting both without changing what counts as a case halfway through.

Two rules do most of the work. Add mutually exclusive routes and multiply successive choices. For overlapping events, add their probabilities and subtract the intersection once. A four-character code made from ten digits has 10 times 10 times 10 times 10 possible strings when repetition is allowed. Forbidding repeated digits changes the count to 10 times 9 times 8 times 7. The probability calculation changes because the sample space changed.

Order is often the hidden question. From ten people, there are 10 times 9 ways to choose a chair and a deputy because the roles differ. There are only 10 times 9 divided by 2 ways to choose an unordered pair. The division removes double counting: Alice with Ben and Ben with Alice are the same pair. Permutations count arrangements. Combinations count selections where order is irrelevant.

Cards make the distinction visible. A five-card hand is an unordered selection from 52 cards, so the number of possible hands is “52 choose 5”, equal to 2,598,960. Dealing the same cards in another order does not create another hand. If the question concerns the sequence in which cards arrive, order returns and so does a larger sample space. Neither count is inherently right. Each answers a different question.

The most efficient count often approaches an event through its complement. The birthday problem asks for the chance that at least two people in a room share a birthday. Counting every matching pattern is awkward. Counting no match is easy. Under the standard model of 365 equally likely birthdays, with leap days ignored and birthdays treated as independent, the first person can have any birthday. The second must avoid it, giving 364/365. The third must avoid two used dates, giving 363/365. Continue multiplying.

With 23 people, the chance of no shared birthday is about 49.27 per cent, so the chance of at least one match is about 50.73 per cent. The result feels too large because people often answer a different question: what is the chance that someone shares my birthday? Among 22 other people, that chance is only about 5.86 per cent. The original question compares every person with every other person. Twenty-three people create 253 pairs, and opportunities accumulate faster than headcount.

This is the general rare-event pattern. A small probability per opportunity can produce a substantial probability across many opportunities. If each of a million independent operations has a one-in-a-million chance of a particular failure, the chance of at least one failure is not one in a million. It is one minus the chance that all million operations avoid it, which is about 63.21 per cent.

Counting can also fail through unequal cases. A total of seven from two dice has six ordered routes: 1 and 6, 2 and 5, through to 6 and 1. A total of two has one. Listing the totals 2 through 12 and assigning each equal probability loses the structure that produced them. The elementary outcomes are the 36 ordered face pairs, not the 11 possible sums.

The deeper habit is to count mechanisms, not labels. Ask how many distinguishable routes reach each result, whether those routes carry equal weight and whether order matters. Once those questions are fixed, many odds that looked mysterious become bookkeeping.

Information Changes the Denominator

Probability changes when information changes because the live possibilities have been restricted or reweighted.

The notation is compact. $P(A \text{ given } B)$ means the probability of event A once B is known. It equals the probability that A and B occur together, divided by the probability of B. Equivalently, $P(A \text{ and } B)$ equals $P(B)$ times $P(A \text{ given } B)$. The denominator is no longer the whole original sample space. It is the part compatible with B.

Take a standard deck. Before seeing anything, the probability that a card is a king is $4/52$. Learn that the card is a face card and the denominator shrinks to the 12 jacks, queens and kings. Four are kings, so the conditional probability is $4/12$. Learn instead that the card is red and the denominator becomes 26 cards, two of them kings, leaving $2/26$. Information matters through what it rules out and how the remaining cases are weighted.

Independence is the special case in which learning B leaves the probability of A unchanged. In symbols, $P(A \text{ given } B)$ equals $P(A)$, which also implies $P(A \text{ and } B)$ equals $P(A)$ times $P(B)$. Two separate fair coin tosses fit the model. Learning that the first was heads does not alter the second toss's one-half chance of heads.

Mutual exclusivity is almost the opposite. If A and B cannot both happen and each has positive probability, learning B makes A impossible. Drawing a king and drawing a queen on one card are mutually exclusive, not independent. Confusing the terms turns ordinary language into bad mathematics: events may sound unrelated while sharing a denominator, or sound opposed while being statistically independent.

The Monty Hall problem exposes the role of information policy. A prize is

behind one of three doors. You choose Door 1. The host knows where the prize is, always opens a different door showing no prize, and always offers you the remaining closed door. Your initial choice wins with probability one-third. The other two doors together carry probability two-thirds. Under the stated protocol, a strategy of switching wins exactly when the initial choice was wrong, because the host must expose the other losing door. It therefore wins two times in three.

Change the host's rule and the answer can change. If a person who does not know the prize location opens a door at random and happens to reveal no prize, the observation carries different information. If the host offers a switch only on selected occasions, the offer itself may reveal something. The famous answer is therefore conditional on the protocol, not on the stage furniture.

Selection rules matter everywhere. Suppose you inspect a sequence only after a striking event appears. The probability of seeing something striking is higher than if you had named that event in advance. Stop an experiment when results look favourable, report only the best subgroup, or search many patterns and the relevant denominator includes all the opportunities to be impressed. That territory belongs mainly to statistics, but the mathematical error is conditional: evidence was evaluated as though the selection process did not exist.

Dependence can also appear after conditioning. Two events may be independent overall but become connected once a third fact is known. If two causes can each trigger an alarm, learning that the alarm sounded can make evidence against one cause count in favour of the other. This kind of induced dependence is why complicated networks cannot be handled by multiplying every visible probability.

Conditional probability is less a formula than a demand for bookkeeping honesty. What was known? How was it learned? Which possibilities remain? Who chose what to reveal? Whenever an answer changes after new information, the denominator is where the change happened.

Bayes Runs the Arrow Backwards

Many probability problems run forwards. Given a cause or model, what is the chance of the evidence? Real decisions often run the other way. Given the evidence, how plausible is the cause?

Bayes' theorem connects the two directions. In its familiar form, the probability of a hypothesis given evidence equals the probability of the evidence given the hypothesis, multiplied by the prior probability of the hypothesis, then divided by the overall probability of the evidence. The denominator normalises the revised weights across the stated hypotheses.

The odds form shows the mechanism more clearly:

posterior odds = prior odds $\times$ likelihood ratio.

Prior odds describe the balance before the new evidence. The likelihood ratio asks how much more probable the evidence would be if one hypothesis were true than if its rival were true. Evidence moves belief to the extent that it discriminates between explanations.

Consider a hypothetical factory producing 100,000 components. Suppose 1 in 1,000 is defective, so about 100 are defective. A scanner flags 99 per cent of defects and falsely flags 1 per cent of sound components. It will flag about 99 defective components. It will also flag about 999 of the 99,900 sound ones. Among roughly 1,098 flags, only about 9 per cent identify a defect.

Nothing went wrong with the scanner. A positive flag is 99 times more likely for a defective component than for a sound one, which is strong evidence. The starting odds were about 1 to 999. Multiplying by 99 moves them to about 99 to 999, or roughly 1 to 10. Strong evidence did not create a high posterior probability because the target was scarce and sound components generated many opportunities for false alarms.

Natural frequencies make the calculation easier to see than percentages stacked inside percentages. Start with a convenient population, apply the base rate, then place true and false flags into separate boxes. The method

does not replace Bayes. It displays the same arithmetic in a form that keeps denominators visible.

Priors provoke argument because they place starting assumptions inside the calculation. Sometimes a prior comes from a known production rate, repeated history or a well-defined physical model. Sometimes it records informed judgement. Sometimes rival priors are reasonable. The honest response is sensitivity analysis: calculate how conclusions change across defensible starting values instead of hiding one choice or declaring that no starting point exists.

Likelihoods need equal care. Evidence is powerful when rival hypotheses predict it differently. A fact that is common under both explains little, however vivid it seems. Several pieces of evidence cannot be multiplied as though independent when they share a source. Ten reports copied from one original witness are not ten independent confirmations. A model that treats them that way manufactures certainty from repetition.

Bayes' theorem is an identity, not an oracle. It cannot supply a missing hypothesis, correct a biased measurement, establish causation or decide which consequences matter. If the true explanation is absent from the model, the available probabilities can still sum perfectly to one. They will divide belief among the wrong options with immaculate consistency.

Its value is narrower and stronger. Bayes separates where belief started from what the evidence contributed. It forces a striking result to compete with the base rate and asks whether the observation would look different under another explanation. Learning then becomes a controlled transfer of weight rather than a swing between certainty and surprise.

A Distribution Is More Than Its Average

An uncertain quantity is represented by a fixed rule assigning a numerical value to each allowed outcome.

That rule is called a random variable. The phrase can mislead because the mapping itself is fixed. What varies is which underlying outcome occurs. For

two dice, the sum is a random variable taking values from 2 to 12. Its distribution assigns probability to those values, with 7 receiving more weight than 2 because more face pairs produce it.

A discrete distribution lists probabilities for separate values. A continuous distribution spreads probability across intervals. Its density may rise above one without violating the rules because density is probability per unit, not probability itself. An exact point can have probability zero while a surrounding interval has positive probability. The cumulative distribution answers a more stable question: what is the probability the value is at or below a chosen threshold?

Different mechanisms produce different families. A fixed number of independent yes-or-no trials with common success probability gives the binomial distribution. In a homogeneous Poisson process, the mean rate is constant and counts in disjoint intervals are independent. Fixed-interval counts are Poisson, while waiting times have an exponential, memoryless distribution. These are models with conditions, not labels attached after seeing a familiar shape.

Expected value compresses a distribution to its weighted centre. Multiply each possible value by its probability and add. The expected result of a fair die is 3.5, although no roll can produce 3.5. Expectation is not the most likely outcome and need not be attainable. It is the balance point of the distribution and, under suitable repetition, the long-run average.

Expectation has a useful property that often surprises people: it is linear even when variables are dependent. The expected value of X plus Y is the expected value of X plus the expected value of Y . This makes totals manageable. Dependence returns when spread matters. The variance of a sum includes covariance terms that measure whether variables tend to move together.

The average alone can conceal everything a decision-maker cares about. Compare two prospects. The first produces £0 with certainty. The second produces a gain of £100 or a loss of £100 with equal probability. Both have

expected value £0. Their experiences are plainly different. A third prospect could share the same expectation while placing tiny probability on a devastating loss. One scalar cannot preserve location, spread, asymmetry and tail shape at once.

Variance measures the expected squared distance from the mean. Squaring prevents positive and negative deviations from cancelling and gives large deviations extra weight. Standard deviation returns the measure to the original units. Both are useful, yet neither tells you which side contains the danger. A skewed distribution may have moderate standard deviation and a long harmful tail.

Quantiles answer questions that averages cannot. The median divides probability in half. A 95th percentile is a threshold below which 95 per cent of the distribution lies. In loss modelling, a high loss quantile can reveal an uncomfortable region, though it still says little about how bad outcomes become beyond that threshold. Tail expectation and scenario analysis may matter more when the final few per cent contain the whole threat.

Shape also affects aggregation. Add many independent, or suitably weakly dependent, contributions with controlled variance and no dominant term, and their centred, scaled sum may approach a normal distribution. Heavy tails, strong dependence, dominant components or changing mechanisms can defeat the approximation where it matters most. A tidy centre does not guarantee a tidy edge.

Probability becomes actionable when the whole distribution is kept long enough to match the question. Expected value prices repeated exposure. Variance describes fluctuation. Quantiles locate thresholds. Tail measures examine extremes. No summary is universally best. Choosing one is another modelling decision, and every summary forgets something.

Repetition Creates Order, Slowly

A coin toss is unstable. A large collection can be remarkably regular. Probability explains both without giving the coin a memory.

For independent trials with common success probability p , the number of successes in n trials is binomial. Its expected value is np and its standard deviation the square root of $np(1-p)$. In 100 fair tosses, 50 heads is the centre, but exact balance is only one possible count. Results several heads away remain ordinary because variation grows with the square root of n .

That square-root rule explains why proportions stabilise. The expected count grows in direct proportion to n , while the typical size of the fluctuation grows more slowly, like the square root of n . Divide by n and the relative wobble shrinks. A law of large numbers formalises the result: under stated conditions, the sample average converges towards the expected value as the number of observations grows.

Convergence does not mean a smooth approach, a prompt one or a guarantee about any finite run. A fair coin can show 60 heads in 100 tosses. It can begin with ten heads. Long streaks are not intrusions into randomness. They are part of it. In a long enough sequence, some pattern that would have looked startling at one named position becomes likely to appear somewhere.

The gambler's fallacy confuses stabilising proportions with compensation. After several heads, the proportion of heads is above one-half. Future tosses can pull the overall proportion towards one-half without tails becoming more likely on the next toss. If each toss is independent, the next probability remains one-half. The past imbalance becomes a smaller fraction of a growing total; the coin does not repay a debt.

The opposite error is to treat independence as the default. Machines wear, weather persists, people learn and populations change. Suppose tomorrow's chance of rain is 70 per cent after a wet day and 20 per cent after a dry one. The sequence remains random but dependent. A two-state Markov chain records this through transition probabilities. Under that model, the current state carries the relevant memory; earlier states add nothing once it is known. Probability supplies both models. The mechanism decides which belongs.

The central limit theorem supplies another form of order. In common versions, centred and scaled sums of independent, or suitably weakly dependent, contributions approach a normal shape when no term dominates and variance is controlled. The contributions need not be normal. This helps explain bell curves in measurement error and aggregated variation. Heavy tails, dominant terms, strong dependence or changing mechanisms can slow or prevent the approximation, especially in the extremes.

Simulation extends repetition from the world into a computer. A pseudo-random generator produces a deterministic sequence designed to behave like draws from a chosen distribution for the intended use. Feed those draws through a model many times and the resulting Monte Carlo sample approximates the distribution of the model's outputs. Complicated integrals, queues, reliability systems and paths can become collections of repeated artificial experiments.

For an ordinary independent Monte Carlo estimate with finite variance, sampling error falls at the familiar square-root pace. To halve typical error, roughly four times as many runs are needed. More runs do nothing to repair a false input distribution, missing dependence or an excluded failure mode. A billion precise simulations of the wrong model remain wrong.

Repetition therefore creates order only under a stable enough structure and only in the quantity the theorem controls. It can stabilise an average while individual paths remain rough. It can produce a smooth simulated distribution while model risk stays untouched. The long run is powerful, but nobody is entitled to reach it.

Risk Begins Where the Average Ends

Probability becomes risk when outcomes acquire consequences. The mathematics can combine chance and loss. It cannot decide how much loss a person, firm or society should accept.

Expected loss is the probability-weighted average of possible losses. It is indispensable for repeated, pooled exposures. An insurer can price many

similar claims around an expected total, while holding capital for variation. A manufacturer can compare expected failure costs across designs. Yet expected loss is not a complete description of risk, because the path, timing and survivability of losses matter.

Two projects can have the same expected cost while one produces frequent small setbacks and the other places a tiny probability on destruction. The second may deserve more attention even when its expected loss is lower. A reversible error differs from an irreversible one. A delayed loss may allow adaptation; an immediate loss may not. A person with ample reserves and a person near insolvency face different decisions under the same distribution.

The St Petersburg game made this gap famous. Toss a fair coin until the first head. Pay £2 if it arrives on the first toss, £4 on the second, £8 on the third, continuing to double. Each possible stage contributes £1 to expected value: probability $1/2$ times £2, probability $1/4$ times £4, and so on. The infinite sum diverges. The game's expected monetary value is infinite, yet no sensible buyer offers an unlimited entry fee.

The paradox does not show that expectation is useless. It shows that money does not have constant personal value across all wealth levels, and that an unbounded mathematical game ignores practical limits on bankroll and payout. Daniel Bernoulli's response was expected utility: represent the value of outcomes to the decision-maker, not merely their cash amount. Later decision theory made that approach systematic. Utility models still require assumptions and do not turn ethics into arithmetic, but they explain why equal expected money need not mean equal choice.

Ruin creates a harder stopping rule. Imagine wealth multiplied by 1.5 after a win and by 0.5 after a loss, each with probability one-half. The expected multiplier is one, so expected wealth remains unchanged from one round to the next. The geometric mean multiplier is the square root of 0.75, about 0.866. Across independent repetitions, a typical long path therefore shrinks while rare high paths hold up the arithmetic mean. A process can look fair in expectation while a typical long path decays.

Dependence changes protection. Ten risks do not diversify much if one cause drives all ten. Fire doors sharing one power supply, suppliers sharing one port or investments sharing one funding shock can fail together. Correlation is one summary, but it does not determine the whole joint distribution. Two systems can share the same correlation while having different chances of simultaneous extreme loss. Combining marginal probabilities under an unjustified independence assumption can make a system look safest where it is weakest.

Exposure changes rare-event language. A one-in-a-million chance per operation says little until the number of operations is known. Time, scale and repeated attempts convert local probabilities into system probabilities. So does adversarial selection: a weakness that random use rarely finds may be discovered quickly by someone searching for it.

Good risk work therefore looks beyond the average. It asks about the full loss distribution, tail scenarios, dependence, available reserves, reversibility and the ability to stop or adapt. It distinguishes uncertainty inside the model from uncertainty about the model. It tests conclusions across plausible assumptions rather than polishing one forecast.

This completes the loop begun with model specification. Probability made chance calculable by defining outcomes and assigning weights. Risk exposes the price of those choices. If the model omits a common cause, a catastrophic branch or the condition under which the process ends, the resulting number can be exact and unusable. The final question is never only “What is the probability?” It is “Probability of what, under which model, across how many exposures, with what consequences if the unlikely branch arrives?”

How It Actually Works

Dividing an unfinished game

In 1494, Luca Pacioli printed an interrupted-game problem. Two players staked 10 ducats in all on a ball game to 60 points, with ten points for each goal. Play stopped with the score at 50 points to 20. How should the stake be divided?

Pacioli's shortest method divided the money in proportion to points already scored, giving five-sevenths to the leader and two-sevenths to the other player. The answer looks even-handed and prices the wrong object. Under the simplifying assumption that either player is equally likely to win each future goal, the leader wins unless the other takes four goals in succession. The fair split is therefore fifteen-sixteenths against one-sixteenth. The past score matters through how many wins each player still needs.

The problem of points mattered because it forced a new kind of arithmetic. The past score was known. The future was not. Fair division required giving present value to continuations that had never occurred.

People in many cultures had long used dice, lots, contracts and numerical reasoning about uncertain life. This chapter traces the surviving line that became modern mathematical probability, not a universal first. Within that line, what was missing was a portable method that treated uncertainty as a quantity. Chance belonged to fortune, providence, judgement or practical craft. It had not yet become a mathematical object whose rules travelled from one problem to another.

Cardano counts the cases

Girolamo Cardano moved unusually close before the seventeenth century. He was a physician, mathematician and experienced gambler whose life supplied more uncertainty than one career required. In *The Book on Games of Chance*, written in stages during the sixteenth century, he analysed dice, cards, cheating, stakes and what a fair wager should mean.

His decisive move was to compare favourable cases with the complete circuit of equally possible cases. With two dice, he recognised that sums have unequal numbers of routes. Nine can arise as 3 and 6, 4 and 5, 5 and 4, or 6 and 3. Two has one route. The labels are not equal because the mechanisms underneath them are not equal.

Cardano also understood the importance of repeated trials and the practical difference between theoretical fairness and a short run. He did not create modern probability. His treatment was incomplete, his terminology unstable and some results wrong. The manuscript was published only in 1663, long after his death, so it did not launch an immediate school. It shows that many ingredients existed before the conventional founding date without yet forming a public discipline.

Pascal, Fermat and the future tree

The usual starting point is 1654, when Blaise Pascal and Pierre de Fermat exchanged letters about games of chance and the problem of points. The correspondence survives unevenly, but the mathematical shift is clear. Instead of dividing the stake by past performance, they counted the possible future sequences needed to finish the game.

Suppose Player A needs one more win and Player B needs two. At most two further rounds are needed. Write the possible two-round sequences as AA, AB, BA and BB, treating each round as fair and independent. A wins the stake in the first three sequences; B wins only in BB. The fair division is therefore three-quarters to A and one-quarter to B.

The sequences AA and AB include a second round that would never be played because A has already won. That does not spoil the calculation. The imagined extra result can be attached without changing who wins the series. Completing the tree to a common depth makes equally weighted paths countable.

Pascal also developed a recursive arrangement of binomial coefficients now called Pascal's triangle, although versions long predated him in several mathematical cultures. The coefficients count how many sequences contain

a given number of successes. Fermat preferred direct enumeration. Their methods met because combination counts and future paths are two views of the same structure.

The importance of the correspondence lies less in the game than in the transfer. An uncertain future could be broken into exhaustive cases, weighted, and converted into a present fair share. Chance had acquired an arithmetic of expectation.

A dice puzzle associated with the same circle shows why counting had to replace intuition. Four throws of one die give a chance of at least one six equal to one minus $(5/6)$ to the fourth power, about 51.77 per cent.

Twenty-four throws of two dice give a chance of at least one double six equal to one minus $(35/36)$ to the twenty-fourth power, about 49.14 per cent. Both experiments have the same expected number of target hits, two-thirds. One crosses 50 per cent and the other does not. Equal expectation does not mean equal probability of at least one success. The complement calculation settles what resemblance cannot.

Huygens gives expectation a textbook

Christiaan Huygens learned of the French work and published *On Reasoning in Games of Chance* in 1657. It was the first printed treatise devoted to mathematical probability and, for decades, the main route by which the new subject travelled.

Huygens began with a principle of fair exchange. If a person has equal chances of receiving a or b, the value of the prospect is halfway between them. More general prospects receive a probability-weighted value. He used this expected value to solve wagers, dice problems and interrupted games, then ended with problems for readers to attack.

Expectation did something conceptually large. It placed unequal futures on one present scale. A claim to £10 with probability one-half could be compared with £5 for certain, at least within a linear monetary model. The move supported fair division and pricing without predicting which future would occur.

The method soon escaped the gaming table. Seventeenth-century governments sold life annuities. Merchants insured voyages. States borrowed against uncertain lives and revenues. John Graunt's 1662 analysis of London mortality bills found regularity in records of deaths. Johan de Witt used survival reasoning to value annuities in 1671. Edmond Halley's 1693 life table, built from records for Breslau, connected ages, survival and annuity pricing more systematically. Probability and data were beginning to meet in institutions that had money exposed to time.

Bernoulli joins chance to frequency

Jacob Bernoulli spent years extending the new calculus. His *Ars Conjectandi*, published in 1713 after his death, included combinatorics, results derived from Huygens and a theorem linking theoretical probability to observed frequency.

The result later became known as the law of large numbers. If independent trials repeat with the same probability, the observed proportion can be made highly likely to lie within a chosen distance of that probability by taking enough trials. Bernoulli proved a version for binary outcomes and understood that the required sample could be enormous.

This was a bridge between two meanings of probability. In games, symmetry can assign a probability before observation. In medicine, law, commerce and politics, the underlying chances may be unknown. Repetition might reveal stable proportions, though Bernoulli knew that social cases were less clean than urns and dice. The theorem did not say that a finite frequency equals a true probability. It explained why frequency could become evidence about one under a stable model.

Bernoulli also widened "conjecturing" beyond games. Decisions had to be made from incomplete signs, testimony and experience. Mathematical probability was becoming a discipline of rational judgement, not merely a way to settle stakes.

de Moivre finds the bell

Abraham de Moivre, a French Protestant exile working in London, published *The Doctrine of Chances* in 1718 and expanded it in later editions. He developed generating functions, recurrence methods and approximations that made larger repeated problems tractable.

In a privately circulated 1733 pamphlet, later incorporated into editions of *The Doctrine of Chances*, he showed that the central part of a binomial distribution could be approximated by a smooth bell-shaped curve. Thousands of discrete counts no longer had to be calculated one by one. A continuous curve could estimate how far a repeated total was likely to wander from its centre.

The curve later became associated with Carl Friedrich Gauss through error theory, and the normal distribution acquired an authority that sometimes exceeded its conditions. Its rise shows a recurring pattern in probability. A method begins as an approximation for a defined mechanism, then becomes a default picture of variation. The approximation is powerful precisely where the aggregation assumptions fit. Outside them, the bell can hide skew, dependence and heavy tails.

Bayes and Laplace reverse the question

Thomas Bayes left an unpublished essay on how to reason from observed outcomes to an unknown chance. After Bayes's death in 1761, Richard Price edited and presented it to the Royal Society. It appeared in 1763.

Bayes considered a stylised experiment in which an unknown probability was itself generated by a random construction. After observing successes and failures, what could be said about the hidden chance? The essay contained the machinery now attached to Bayes's name, although the modern theorem, notation and range of use emerged through later development.

Pierre-Simon Laplace made the reverse direction central. From the 1770s onward, and at full scale in his 1812 *Théorie analytique des probabilités*, he

developed methods for moving from effects to causes, combining observations, approximating distributions and treating errors. The elementary identity connecting $P(A \text{ given } B)$ and $P(B \text{ given } A)$ became part of a broad programme of inference.

Laplace's work also displayed the ambition of classical probability. If the state of the universe and every force were known, a perfect intelligence could calculate future and past. Human probability reflected incomplete knowledge. That determinist picture did not prevent chance mathematics from expanding. It gave probability a role as the calculus of ignorance.

Randomness starts moving

Early probability problems asked which result would occur after a fixed number of trials. Nineteenth-century science and finance needed paths: quantities wandering through time, with the next position built from earlier movement. The object of study was no longer a final count but an entire uncertain trajectory.

In 1900, Louis Bachelier's doctoral thesis, *Theory of Speculation*, modelled price changes through a continuous random process. His financial assumptions were idealised and markets do not follow them cleanly, but the work is recognised as an early Brownian-motion model. Five years later Albert Einstein used a related diffusion model for particles suspended in liquid, connecting visible irregular motion to molecular activity. Jean Perrin's measurements then supported the molecular account and helped make Brownian motion evidence for atomism.

The mathematical difficulty was severe. A Brownian path is continuous yet jagged at every scale, and there are infinitely many possible paths. Norbert Wiener placed a rigorous probability measure on such path space in the early 1920s. Probability could now assign weight to events involving whole histories, such as whether a moving quantity crosses a boundary before a deadline.

This shift made stochastic processes central. A probability distribution could evolve through time, and questions about hitting, waiting, persistence and

maximum loss became mathematical objects. The same language could describe diffusion in physics while remaining only a model when applied elsewhere. Shared mathematics did not imply shared causes.

Probability acquires foundations

During the nineteenth century, probability entered astronomy, error measurement, insurance, demography, physics and social statistics. Its successes sharpened an old question: what kind of thing was a probability?

Classical probability relied on equally possible cases, but “equally possible” could sound circular when probability was used to define it. Frequency interpretations tied probability to long-run proportions, yet an infinite long run is not observed. Subjective and logical approaches treated probability as rational support or coherent belief, raising questions about whose information and which rules of judgement should govern.

The mathematics also outgrew finite counting. Random motion, continuous time and infinitely many possible paths required a language capable of assigning probability consistently to complicated sets. Work on measure by Émile Borel and Henri Lebesgue supplied much of the machinery.

In 1900, David Hilbert called for an axiomatic treatment of probability alongside parts of physics. Andrey Kolmogorov delivered the decisive synthesis in 1933 with *Foundations of the Theory of Probability*. He treated probability as a measure on a collection of events. Non-negativity, total probability one and countable additivity became the formal base from which conditional probability, random variables and distributions could be built.

The axioms settled how valid probabilities behave. They did not settle whether a probability is physical chance, long-run frequency, rational belief or something else. That division of labour was productive. Mathematicians could prove theorems inside the framework while scientists and philosophers argued about how models connect to the world.

Dependence gets its own machinery

The earliest games encouraged a clean picture of repeated independent trials. Many processes remember where they have been. Tomorrow's weather depends on today's. A queue length affects the next waiting time. A molecule's next position begins at its current one. Treating such systems as fresh coin tosses loses the mechanism.

Andrey Markov developed a framework in which the distribution of the next state depends on the current state, while earlier history matters only through that present state. This memory rule now defines a Markov chain. The chain moves among states according to transition probabilities, so repeated multiplication describes how an initial distribution spreads, settles or keeps circulating.

In 1913 Markov chose an arresting demonstration. He classified the first 20,000 letters of Pushkin's *Eugene Onegin* as vowels or consonants and counted how often one category followed the other. There were 8,638 vowels and 11,362 consonants, but the sequence was not independent. The chance that the next letter was a vowel differed according to whether the previous letter was a vowel or consonant. Literature became evidence that dependent trials could still be analysed mathematically.

The importance reaches beyond the anecdote. A chain can be random without being independent from step to step. It can forget its starting point gradually, become trapped, revisit states or possess a long-run distribution. Random walks, queues, population models and sampling algorithms all use versions of this idea. Probability had moved from counting isolated outcomes to describing motion through a structured set of possibilities.

Chance enters the computer

By the twentieth century, many probability models were too complicated for closed-form calculation. Computation offered another route: imitate the random process repeatedly and study the resulting sample.

At Los Alamos after the Second World War, Stanislaw Ulam, John von

Neumann, Nicholas Metropolis and colleagues developed the modern Monte Carlo method while working on neutron transport and other problems. Ulam later described thinking about repeated random trials while playing solitaire during illness. The method drew artificial random inputs, passed them through a model and used frequencies or averages to estimate quantities that were difficult to integrate directly.

The name referred to Monte Carlo's casinos. The work was less glamorous. The first computerised Monte Carlo calculations ran on ENIAC in April and May 1948, with later phases following. Metropolis and Ulam published a general account in 1949.

Computers are deterministic machines, so most simulations use pseudo-random sequences generated by algorithms. Good generators are tested for the patterns that would corrupt their intended applications. Security requires stronger unpredictability than routine simulation. A sequence adequate for estimating an integral may be disastrous for cryptographic keys.

Monte Carlo changed the practical reach of probability. Reliability models, queueing systems, particle paths, financial scenarios and Bayesian calculations could be explored through repeated artificial trials. Later methods learned to sample from difficult distributions and update large systems. The logic remained the old one: specify outcomes, assign weights, condition on information and aggregate consequences. The computer increased the number of cases that could be explored. It did not certify that the right model had been built.

Simulation also changed what counted as an answer. A formula can reveal structure across every parameter value. A simulation often gives an estimate for one chosen set of inputs, accompanied by numerical uncertainty. The trade is justified when the formula is unavailable, but it requires diagnostics. Independent reruns, convergence checks and alternative generators test the computation. Changed assumptions test the model. Confusing those checks produces confidence in digits while the event set or mechanism remains wrong.

How we know

The mathematical claims in this book differ from historical and applied ones. A theorem follows from definitions and assumptions, and its proof can be checked line by line. Whether a die, forecast, test or financial model fits those assumptions is an empirical question that proof cannot answer.

The early history survives through printed treatises, letters and later editions. Pacioli's original problem specifies a 10-ducat ball game stopped at 50 to 20, and his proportional division can be checked against the text. Cardano's manuscript predates its publication and had limited immediate reach. The Pascal-Fermat exchange is a strong landmark because it influenced Huygens and later writers, but it was not humanity's first encounter with quantified uncertainty. "First" claims are therefore narrowed to a surviving publication, proof or line of transmission.

Bayes's essay reached print through Richard Price, while Laplace supplied much of the later generalisation. Kolmogorov's 1933 monograph establishes the axiomatic framework directly. Archival work fixes the first computerised Monte Carlo calculations to ENIAC in April and May 1948, although participant recollections still compress collaboration.

For applications, the strongest test is whether assumptions, calibration, dependence and failures survive relevant data and changed conditions.

What People Get Wrong

"A probability predicts the next outcome"

A 70 per cent probability is often heard as a weak promise that the event will happen. When it does not, the number is declared wrong. The mistake comes from forcing a statement about uncertainty into the grammar of a single outcome.

One realised outcome can contribute evidence, especially when rival forecasts assigned it different probabilities, but it cannot reveal a calibration rate by itself. An event given 70 per cent may occur or fail. Calibration is

assessed across forecasts made under a defined procedure: among cases assigned around 70 per cent, the event should occur around seven times in ten. A system can be calibrated yet uninformative if it rarely departs from the base rate, so resolution and proper scoring matter too.

That distinction matters because hindsight turns uncertainty into a verdict. A low-probability outcome does not prove negligence, and a high-probability outcome does not prove skill. Decisions should be judged against the distribution visible beforehand, then updated with what the result teaches about the model. A forecast can miss on Tuesday and remain defensible; a forecasting system can be right once and remain worthless.

“Random means evenly spread”

People expect a random sequence to alternate politely and cover its possibilities at regular intervals. Genuine randomness is rougher. It clusters, leaves gaps and produces streaks. A shuffled deck can contain adjacent cards of the same rank. A fair coin can show six heads in a row. Those patterns are unlikely at a named location but unsurprising somewhere in a long sequence.

Uniform and random are different ideas. Uniform describes equal probability across specified outcomes. Random describes generation according to some probability model. A weighted die can be random without being uniform. A perfectly alternating sequence of heads and tails is evenly balanced but suspiciously non-random because its local structure is too orderly.

The misconception survives because humans judge randomness by appearance and underproduce runs when inventing sequences. In auditing, simulation and everyday inference, clumping alone does not establish manipulation. The right question is whether the pattern is improbable under the complete model, including all the places where a striking pattern could have appeared. Tests of randomness therefore inspect many forms of structure rather than rewarding visual mess.

“After a run, the opposite is due”

Five heads in a row make tails feel overdue. Under independent fair tosses, the next probability remains one-half. The law of large numbers does not force the sequence to repair its balance. Future observations dilute the existing surplus as the denominator grows.

The intuition contains a fragment of truth placed in the wrong location. Across many future tosses, the overall proportion is likely to move closer to one-half because the fixed run becomes a smaller part of the total. That says nothing about compensation on the next trial.

The correction is conditional. If the mechanism can change, a run may carry information. A machine may be drifting, a player may be learning, or weather may persist. Then independence is the claim to test rather than the answer to assume. This is why both the gambler's fallacy and a reflexive dismissal of every streak can fail. Ask whether trials share a stable mechanism and whether past outcomes alter the distribution of the next. The calculation follows the answer; it cannot supply it.

“Independent means unrelated”

In ordinary speech, independent things have no connection. In probability, independence means that learning one event does not change the assigned probability of another, within a specified model. That is narrower and more technical.

Two measurements can be caused by separate mechanisms yet become dependent after selection. Two symptoms may be independent among all people but associated within a group chosen because at least one symptom is present. Conversely, variables linked by a deterministic-looking story can be independent under a distribution that balances their cases in a particular way.

Pairwise independence also need not imply mutual independence. Toss two fair coins and define a third result as whether the first two match. Each individual result is equally likely, and any pair is independent. Once the two

coin results are known, however, the match result is fixed. Every pair can pass an independence check while the three together remain constrained. Multiplying all probabilities would then fail.

Independence matters because it permits products, diversification and simple repeated-trial models. It must be justified at the level used. Naming separate objects, departments or safeguards does not prove their failures are independent when they share power, weather, data or incentives. Independence can hold conditionally within each environment and disappear after environments are mixed, so the reference population matters too.

“A one-in-a-million event is negligible”

A small probability has no practical meaning until exposure is known. One chance in a million per operation can be remote for one operation and likely somewhere across a million independent operations. The chance of at least one occurrence is then about 63 per cent.

The same logic drives coincidences and false alarms. Search enough dates, symptoms, transactions, words or subgroups and an individually unusual match becomes likely to appear. Reporting the probability of the selected match as though it had been named in advance ignores the search that found it.

Dependence can move the answer in either direction. Repeated exposures sharing one cause do not behave like independent trials. An attacker choosing where to probe is not random exposure. Time can change the underlying rate. The probability attached to one unit cannot be scaled safely until the mechanism connecting units is known, and the number of opportunities is stated explicitly and clearly.

Without the denominator, “one in a million” often functions as reassurance rather than analysis. Ask per what, for how long, across how many units, under which dependence, and whether the event was specified before or after the search. A rare cause may be common among selected disasters even while remaining rare across all ordinary cases.

“The average tells you what to expect”

Expected value is often called the expected outcome, inviting the idea that it describes a typical experience. A fair die has expected value 3.5, which cannot be rolled. A highly skewed distribution can have a mean far above what most cases receive. Two prospects can share a mean while having opposite tail risks.

Expectation is a weighted centre. It is powerful for totals across repeated or pooled cases and for comparing prices under suitable assumptions. It does not preserve variance, quantiles, timing, dependence or survivability. Nor does it decide whether money has the same value to people with different resources.

The word “expect” is therefore dangerous only when left unqualified. Ask whether the mean, median, mode, range, downside quantile or chance of ruin matches the decision. The distinction becomes decisive when one outcome can end the process. A favourable average cannot be collected by someone who cannot survive the path required to reach it. Even without ruin, an average across people may describe nobody's repeated experience.

“More decimal places mean less uncertainty”

A probability reported as 0.7314 looks more authoritative than one reported as about 70 per cent. The extra digits may reflect calculation, not knowledge. A simulation can estimate the output of a chosen model to four decimal places while the input rate is uncertain by ten percentage points.

There are at least two kinds of uncertainty. Sampling or numerical error concerns how precisely the model's answer has been computed. Model uncertainty concerns whether the possibilities, weights, dependence and mechanisms are right. More data or more simulation can reduce the first while leaving the second untouched.

False precision is persuasive because arithmetic is visible and assumptions are not. Rounding can also conceal useful distinctions, so the answer is not

compulsory vagueness. Match the reported precision to the weakest material input, show ranges where assumptions matter and separate measured variation from judgement. A wide interval supported by a sound model can carry more information than a narrow number built on an omitted branch. Precision should be earned by the whole chain, not borrowed from the calculator at its end.

Use It

Name the event and the horizon

A probability should make you ask for a complete sentence. Chance of what, by when, for whom, under which conditions?

“Failure risk is 2 per cent” could mean 2 per cent per component, per year, per journey or across the life of the system. It could count any interruption or only permanent loss. The same number can describe radically different exposure once the event and horizon are changed.

Write the event so that two observers could agree whether it happened. Add the time window and unit of exposure. Then ask whether repeated windows overlap or share causes. A monthly probability cannot be multiplied by twelve unless the model supports that approximation. A per-user rate cannot be turned into a system rate without knowing how users interact.

This discipline is especially useful when a claim sounds reassuring. A small number without a horizon is an adjective wearing mathematical clothes. Naming the event converts it into something that can be counted, conditioned and challenged. It also prevents quiet substitution, where evidence about minor incidents is used to reassure about catastrophic failure, or annual risk is quoted beside a lifetime decision.

Build the denominator before reading the percentage

Percentages arrive with a numerator in the spotlight and a denominator in the dark. Reverse the order.

Suppose a screening system is described as catching 99 per cent of defective items. Before treating a flag as decisive, ask how many items are defective, how often sound items are flagged, and which population entered the system. Put 100,000 imagined items on the page. If 100 are defective, 99 may be caught. A 1 per cent false-positive rate among 99,900 sound items produces about 999 false flags. The positive result now means about 9 per cent, not 99 per cent.

Natural frequencies are often easier than nested percentages because they preserve the groups being compared. They also reveal when the base rate came from the wrong population. A rate among referred cases cannot be applied unchanged to everyone.

The habit is portable: draw boxes, place counts inside them, and ensure every percentage has a named reference class. Most base-rate mistakes become visible before Bayes' formula is written.

Compare how well rival explanations predict the evidence

Evidence matters through discrimination. Ask not merely whether an observation fits your preferred explanation, but whether it fits a serious rival almost as well.

A system failure after heavy rain is compatible with water damage. It may also be common when maintenance is poor, regardless of weather. The likelihood ratio asks how much more expected the evidence is under one account than another. A vivid observation that both accounts predict carries little resolving power.

This lens helps separate surprise from support. An unusual event under the status quo can count against it, but only if the alternative made the event less unusual. "I could not imagine this happening by chance" is incomplete until chance is defined and competing mechanisms are modelled.

Treat copied evidence cautiously. Five dashboards drawing from one database are one underlying source. Several indicators responding to the

same shock are dependent. Multiplying them as independent can inflate confidence. Build an evidence tree that shows shared causes before combining weights.

Carry the distribution into the decision

Do not stop at the mean. Ask what outcomes are possible, where the mass sits, how wide the spread is and what lives in the tail.

For a repeated, pooled exposure, expected value may be the right starting point. For a one-off decision, a downside quantile, worst credible case or chance of crossing a survival threshold may dominate. A project with positive expected value can be a poor choice when one branch exhausts the cash needed for every later opportunity.

Separate probability from consequence. A common minor loss and a remote irreversible loss require different controls. Then add reversibility and response time. A bad outcome detected early may be corrected; the same loss arriving without warning may not.

Finally, ask whose utility is being represented. A £10,000 loss does not have one universal weight. A firm with deep reserves, a household near its limit and a public institution responsible for others face different constraints. The distribution can be shared while the decision differs.

Stress the model, then simulate

A model should be attacked before it is polished. Change the assumptions that carry the conclusion: base rate, dependence, tail thickness, exposure, recovery time and the rule for stopping. If the decision flips under a small plausible change, the sensitivity is part of the answer.

Simulation helps when many uncertain parts interact. Generate inputs from the stated distributions, preserve their dependence, run the mechanism and inspect the output distribution. Repeat enough times to reduce numerical noise, then rerun under rival assumptions. The first exercise tests calculation. The second tests the model.

Look especially for failure modes excluded by convenience. Models often assume components fail independently because multiplication is easy, demand remains stable because history is available, and losses stop at the edge of the dataset because nothing beyond it was recorded. Add common shocks and structural breaks as scenarios even when their probabilities are hard to estimate.

Simulation should widen attention, not decorate one forecast. Its most useful output may be the discovery that the answer depends on an assumption nobody had noticed.

The limits

Probability cannot create information that is absent. It can distribute belief across stated possibilities, but it cannot guarantee that the true mechanism appears among them. A complete set of probabilities can sum to one inside a model that has omitted the event that matters.

It also cannot turn every uncertainty into a stable frequency. New technologies, political breaks, rare disasters and strategic opponents may provide little relevant repetition. Subjective probabilities can still organise thought, but their apparent exactness should not exceed the evidence. Sometimes a range, scenario set or refusal to quantify carries more honesty than a single number.

Causal questions require more than association. Conditional probability can describe how variables move together and Bayes can update among causal models. Neither proves which intervention will change the outcome. That requires design, mechanism and evidence beyond probability's internal rules.

Probability cannot decide values. Expected utility represents preferences under assumptions; it does not establish whose welfare counts, whether a risk may be imposed on someone else or which losses are unacceptable. A low expected cost can conceal an unfair distribution of harm. Mathematics can show the trade. It cannot supply consent.

Finally, long-run theorems do not promise any person a long run. Institutions fail, bankrolls end and time runs out. The average may stabilise after the decision-maker has disappeared from the process. Survival conditions belong inside the model from the start.

The one thing to keep

Keep the missing clause.

Every probability statement carries conditions, even when fluent speech erases them. “There is a 20 per cent chance” means 20 per cent for a defined event, under a particular model, given particular information and across a particular horizon. The number is the end of that sentence, not the beginning.

Once you hear the missing clause, familiar claims change shape. A rare event invites a question about exposure. A positive test invites a question about base rates and false alarms. An average invites a look at the distribution. A collection of safeguards invites a question about common causes. A precise simulation invites a question about the mechanisms and branches it omitted.

This does not make probabilistic judgement timid. It makes confidence conditional in the way the mathematics always was. You can still decide firmly. You can still prefer one model, place a bet on one future or accept a risk. The difference is that the discarded branches remain visible, and an outcome cannot rewrite what was knowable before it arrived.

Probability is the discipline of representing several possible outcomes without pretending that they are equally likely. Good judgement adds the final demand: never let the number hide who defined the event, what mechanism or evidence supplied the weights, how often the exposure repeats, and what happens if the thin tail becomes the outcome you must live through.

Terms

Outcome

One possible result represented by a probability model. What counts as one outcome depends on the question: a die face, a path, a waiting time or a loss. The choice controls every later count.

Sample space

The complete set of outcomes allowed by a model, often written as Ω . Missing possibilities create model error before any probability calculation begins. A space may be finite, countable or continuous.

Event

A set of outcomes sharing a property of interest, such as rolling an even number. Events can overlap, contain one another or cover the whole sample space. Probabilities attach to these sets rather than to words alone.

Probability

A non-negative weight assigned to an event, with total weight one across the sample space. Its interpretation depends on the model and information used.

Odds

Relative weight for an event against its complement. To convert odds a to b , divide a by the total $a + b$; probability p gives odds p to $1 - p$.

Complement

The event that a stated event does not occur. Its probability is one minus the original probability, often making “at least one” calculations easier.

Union and intersection

The union contains outcomes in A or B or both. The intersection contains outcomes in both A and B. Their distinction controls addition and conditioning.

Mutually exclusive

Events that cannot occur together, so their intersection is empty. Positive-probability mutually exclusive events are dependent because observing one rules the other out.

Conditional probability

The probability of A after B is known, written $P(A \text{ given } B)$. It restricts attention to outcomes compatible with B and then reweights that smaller space. The method of learning B can carry information too.

Independence

A relationship in which learning one event does not change the probability of another. It permits multiplication, but must be justified within the chosen model. Pairwise independence need not extend to a whole collection.

Bayes' theorem

The identity connecting $P(A \text{ given } B)$ with $P(B \text{ given } A)$, base rates and total evidence. It formalises how probability weights change when information arrives.

Prior and posterior

The prior is the probability assigned before specified evidence is included. The posterior is the probability after it has been incorporated. Both remain conditional on the hypotheses, model structure and information-quality assumptions. Sensitivity analysis shows how much the starting weights matter.

Likelihood

The probability or density of observed evidence under a specified hypothesis, viewed as a function of that hypothesis. It need not sum to one across hypotheses.

Likelihood ratio

The probability of evidence under one hypothesis divided by its probability under another. It measures how strongly the observation discriminates between those two explanations.

Random variable

A numerical mapping defined on the sample space, such as the sum of two dice or total loss. Randomness enters through which outcome occurs.

Probability distribution

The allocation of probability across a random variable's possible values. It preserves information a single summary discards. Shape, spread and tails all belong to it.

Mass and density

A probability mass function assigns probability to each value of a discrete variable and sums to one. A probability density describes a continuous variable: area over an interval gives probability. Density is not itself probability and can exceed one when units are narrow.

Expected value

The probability-weighted average of a random variable. It is a centre of mass and long-run average under conditions, not necessarily a possible or typical outcome. Linearity does not require independence.

Variance and standard deviation

Variance is the expected squared distance from the mean, giving large deviations extra weight. Standard deviation is its square root, returned to the variable's units. Neither identifies skew, tail direction or worst credible loss.

Covariance and correlation

Covariance measures whether two variables tend to deviate from their means together. Correlation rescales it to a dimensionless number between -1 and 1 when variances are positive and finite. Neither captures every form of dependence or tail co-movement.

Quantile

A threshold below which a chosen proportion of probability lies. The median is the 50th percentile; high loss quantiles locate, but do not describe, the far tail.

Bernoulli trial

A trial with two coded outcomes, commonly success and failure, and success probability p . Repeated independent Bernoulli trials produce the binomial model.

Binomial distribution

The distribution of the number of successes in a fixed number of independent Bernoulli trials with common probability p . Its mean is np .

Normal distribution

A symmetric bell-shaped continuous distribution determined by mean and variance. It often approximates aggregated variation under central-limit conditions, but is not a universal law and can understate skew or tail risk.

Poisson distribution

A model for fixed-interval counts in a Poisson process, where counts in disjoint intervals are independent and the rate is stable. It is useful for rare counts, but clustering, seasonality or changing exposure can invalidate it.

Exponential distribution

The waiting-time partner of a Poisson process. It is memoryless: conditional on having waited already, the remaining distribution is unchanged. Real waiting systems often remember age, queues or maintenance.

Law of large numbers

A family of theorems showing that sample averages converge towards expected values under stated conditions. It does not require short-run balance or compensation after streaks. Convergence may be slow along finite paths.

Central limit theorem

A family of results under which centred and scaled sums approach a normal distribution. Dependence, dominant terms or heavy tails can defeat common versions.

Markov chain

A process moving among states by transition probabilities, with the next-state distribution depending on the current state. It may be dependent between steps despite remaining random. Chains can settle, become trapped or forget their starting state.

Monte Carlo simulation

Approximation by repeated random or pseudo-random sampling from a model. Under finite-variance sampling, more runs reduce numerical error; wrong assumptions and missing mechanisms remain. Computation must be separated from model adequacy.

Go Deeper

The broad intellectual tour

Persi Diaconis and Brian Skyrms, *Ten Great Ideas about Chance* (Princeton University Press, 2018). Begin here to see probability as a set of arguments rather than a formula sheet. A mathematician and a philosopher connect early games, Bayes, frequency, axioms and psychological error without pretending that Kolmogorov's rules settled what probability means. Some passages assume comfort with elementary mathematics, so read slowly where the notation thickens. Its sharpest discussion concerns the unresolved tension between physical chance, observed frequency and coherent personal belief. The chapters can be read selectively because each rebuilds enough context for its central argument.

The history of the idea

Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*, second edition (Cambridge University Press, 2006). Hacking asks why Europe could gamble for centuries without producing a general concept of probability, then reconstructs the conditions under which chance became both measurable frequency and reasonable belief. It is a major interpretation rather than a neutral chronology, and later historians have challenged parts of its boundary dates. That tension makes it more useful, not less, once the basic story is in place. Read it beside the surviving treatises rather than as the last word on who invented what.

The mathematical classic

William Feller, *An Introduction to Probability Theory and Its Applications*, Volume I, third edition (John Wiley & Sons, 1968). Feller is where elegant examples become serious probability. Random walks, ruin, recurrence, generating functions, limit theorems and distributions are developed with proofs and an unusual amount of personality. It is not a beginner's leisure read. Work with paper, accept that some exercises will resist you, and use a modern introductory text when notation or omitted steps become a barrier. The reward is contact with the subject's internal machinery and with examples, especially random walks and ruin, that expose what smooth summaries leave out.

Risk in usable form

Gerd Gigerenzer, *Calculated Risks: How to Know When Numbers Deceive You* (Simon & Schuster, 2002). Gigerenzer's strongest practical lesson is that many conditional probabilities become clearer when translated into natural frequencies. The book applies that method to testing, screening and public risk communication, showing how percentages can conceal the reference classes that produce them. Some examples and institutional contexts are now dated, and the argument occasionally presses one representation harder than every task warrants. The central technique remains valuable: rebuild the denominator before trusting the headline number. It carries Bayes' theorem from the page into decisions made under pressure.

Notes and Sources

The Whole Thing in One Page and Why You Should Care

Probability statements and calibration. The manuscript treats a probability as attached to a defined event within a model and information state, while leaving physical, frequency and epistemic interpretations open. Forecast calibration means that, across a defined collection of cases assigned probability p , the event occurs at roughly rate p . One realised outcome can affect a comparative score, but it cannot establish a calibration rate by itself. Glenn Brier's 1950 paper supplied an early proper scoring rule; modern assessment also examines resolution, discrimination and changing base rates.

Probability and statistics. The distinction in the text is directional rather than absolute. Probability commonly moves from a specified model towards implications for outcomes. Statistics commonly moves from observed data towards estimates, tests or comparisons among models. Modern work repeatedly travels in both directions, and Bayesian inference joins them explicitly.

Recomputed examples. All numerical examples were recalculated independently. Under the simplified birthday model, the probability of at least one shared birthday among 23 people is

$$1 - (365 \times 364 \times \dots \times 343) / 365^{23} = 0.507297\dots$$

The chance that at least one of 22 other people shares one specified person's birthday is

$$1 - (364/365)^{22} = 0.058571\dots$$

For one million independent exposures, each with failure probability one in one million, the probability of at least one failure is

$$1 - (1 - 10^{-6})^{1,000,000} = 0.632121\dots$$

These are model calculations, not universal empirical frequencies. Real birthdays are not uniform and can be related within families; operational failures may be dependent.

Core Ideas support

Models, events, axioms and interpretations. Kolmogorov's 1933 formulation treats probability as a measure on events. The book states the rules informally: non-negativity, total mass one and countable additivity across mutually exclusive events. William Feller and Patrick Billingsley support the elementary development. The distinction among physical probability, stable frequency, evidential support and coherent degree of belief follows Ian Hacking, Persi Diaconis and Brian Skyrms, and the current Stanford Encyclopedia of Philosophy survey. The manuscript does not reduce objective chance to ignorance or treat one interpretation as settled by the axioms.

Bertrand's chord problem. Joseph Bertrand used the random-chord example in *Calcul des probabilités* (1889) to show that the answer depends on the mechanism used to select a chord. The values one-third, one-quarter and one-half correspond to three different sampling procedures. E. T. Jaynes later used invariance arguments to ask which procedure makes a problem well posed. The manuscript uses the paradox only for its secure lesson: the phrase “random chord” does not define one probability experiment.

Counting and combinations. The number of five-card subsets of a 52-card deck is $52 \text{ choose } 5 = 2,598,960$. The birthday calculation uses complements and assumes 365 equiprobable, independent birthdays. Persi Diaconis and Frederick Mosteller discuss coincidences and the modelling choices behind such examples. The one-in-a-million calculation uses the complement formula under independence.

Conditional probability and Monty Hall. Steve Selvin's 1975 letters gave the problem its modern published form. The two-thirds success rate for an always-switch strategy requires a host who knows the prize location, always

opens an unchosen losing door and always offers the switch. Change the host's knowledge, selection rule or offer policy and the conditional probability may change. The manuscript states the protocol before giving the answer.

Bayes and the base-rate example. The factory scanner is hypothetical. In a population of 100,000, a defect prevalence of 0.1 per cent gives 100 defective and 99,900 sound components. With sensitivity 99 per cent and false-positive rate 1 per cent, expected flags are 99 true positives and 999 false positives. The positive predictive value is $99 / 1,098 = 0.09016\dots$, about 9 per cent. The likelihood ratio for a positive result is $0.99 / 0.01 = 99$. Natural-frequency presentation follows Gerd Gigerenzer's practical treatment. No diagnostic or manufacturing claim is inferred from the invented parameters.

Priors and evidence. Bayes' theorem is a mathematical identity once the joint model is fixed. Its application still depends on the prior, likelihood, hypotheses and data-generating process. Sensitivity analysis is therefore part of the result when defensible priors differ. The warning about copied reports concerns dependence: several observations sharing one source cannot be multiplied as independent confirmations.

Distributions, expectation and spread. Expected value is the probability-weighted mean when it exists. Its linearity does not require independence. Variance of a sum does depend on covariance. Quantiles locate thresholds but do not describe losses beyond them. The normal approximation is deliberately conditional: common central limit theorems require controlled variance, no dominating term and independence or suitably weak dependence. Feller and Billingsley support the mathematical account; Alexander McNeil, Rüdiger Frey and Paul Embrechts support the distinction among means, quantiles, tail measures and dependence in risk work.

Law of large numbers and central limit theorem. Jacob Bernoulli proved an early law of large numbers for repeated binary trials, published posthumously in 1713. Eugene Seneta traces the theorem's later

development and changing forms. The manuscript does not claim that finite sequences converge monotonically or that trials compensate for earlier deviations. The central limit theorem is described as a family of results rather than one condition-free law.

Dependence and Markov chains. The text rejects automatic compensation after a streak and the opposite assumption that every sequence is independent. A two-state weather example is explicitly hypothetical. In a Markov chain, the next-state distribution depends on the present state, while earlier history adds no further information once that state is known under the model. Markov's 1913 analysis of *Eugene Onegin* supplies the historical anchor. The memory rule is a modelling claim, not a universal property of dependent processes.

Monte Carlo error. For ordinary independent Monte Carlo estimates with finite variance, standard error commonly falls at the inverse square root of the number of draws. Halving that error therefore requires about four times as many draws. More computation reduces numerical sampling error inside a fixed model. It does not correct poor inputs, missing dependence or omitted mechanisms. Nicholas Metropolis and Stanislaw Ulam's 1949 paper gives the general method; Christian Robert and George Casella provide the later statistical development.

The St Petersburg game and utility. The stated game pays $£2^n$ when the first head occurs on toss n . Its expected monetary value is the sum over n of $2^{-n} \times 2^n$, so every term contributes $£1$ and the series diverges. Daniel Bernoulli's 1738 paper, available in Louise Sommer's 1954 translation, proposed diminishing marginal utility as a response. The manuscript also records the model's unbounded payout and practical resource limits, which matter separately from utility curvature.

Multiplicative risk. With independent equal chances of multiplying wealth by 1.5 or 0.5, the arithmetic expected multiplier is 1. The geometric mean multiplier is $\sqrt{0.75}$, about 0.866. Under repeated independent play, the logarithmic growth rate is negative, so a typical long path shrinks while rare high paths sustain the arithmetic mean. The example is mathematical and

does not prescribe one universal investment criterion.

Dependence and tails. Variance, covariance and correlation describe parts of joint behaviour but do not fully determine extreme co-movement. Risk aggregation therefore depends on the joint model, common causes and tail assumptions. The text avoids treating correlation as a complete measure of dependence or a stable crisis parameter.

Historical and operating support

Pacioli and the problem of points. Pacioli's 1494 *Summa de arithmetica* states a ball game to 60 points, ten points per goal, with 10 ducats staked and play stopped at 50 to 20. His shortest method divides in the score ratio, five-sevenths to two-sevenths. Under the body's explicit equal-and-independent-goal model, the leader loses only if the other player wins four successive goals, giving probabilities fifteen-sixteenths and one-sixteenth. Anders Hald and A. W. F. Edwards trace the later development. The manuscript does not claim Pacioli invented the question or that equal player strength was part of his text.

Cardano. Girolamo Cardano's *Liber de Ludo Aleae* was written in stages during the sixteenth century and printed in 1663. It contains favourable-case reasoning, dice enumeration, practical gambling discussion and errors. Sydney Henry Gould's translation and David Bellhouse's close analysis support the account. Because the work circulated poorly before publication, it is evidence of early ingredients rather than a simple founding event.

Pascal and Fermat. The 1654 exchange concerns the problem of points and related dice questions. The surviving record is incomplete and weighted towards Pascal's side. The future-path example in the manuscript is a standard reconstruction of their method, not a quotation from a surviving letter. The historical significance lies in the calculable distribution of an unfinished stake and in the line of transmission to Christiaan Huygens. Edwards and Hald are the principal guides used.

The de Méré calculation. Four throws of one fair die produce at least one six with probability $1 - (5/6)^4 = 0.517747\dots$. Twenty-four throws of two fair

dice produce at least one double six with probability $1 - (35/36)^{24} = 0.491404....$ Both counts have expected target hits $2/3$, showing that equal expectation does not imply equal chance of at least one hit. The association with Antoine Gombaud, chevalier de Méré, is conventional in the history and does not establish that one exchange began from one clearly documented puzzle.

Huygens. *De Ratiociniis in Ludo Aleae* appeared in 1657 as part of Frans van Schooten's *Exercitationum mathematicarum libri quinque*. It is generally described as the first printed treatise devoted to mathematical probability and became the main early textbook of the field. Huygens developed expectation from fair exchange and included problems that stimulated later work.

Mortality and annuities. John Graunt's 1662 analysis of London Bills of Mortality identified regularities in population records. Johan de Witt's 1671 report applied survival reasoning to life annuities. Edmond Halley's 1693 paper constructed a life table from Breslau records and used it to value annuities by age. The text presents these as steps in joining probability, data and financial institutions, not as one sole origin of actuarial science.

Bernoulli and de Moivre. Jacob Bernoulli's *Ars Conjectandi* was published in 1713, eight years after his death. Its theorem shows, for repeated trials under defined conditions, that observed proportions can be made highly likely to lie near the underlying chance. Abraham de Moivre's *Doctrine of Chances* first appeared in English in 1718. His privately circulated 1733 *Approximatio* connected the binomial distribution's central region to the bell-shaped curve and was incorporated into later editions.

Bayes and Laplace. Thomas Bayes died in 1761. Richard Price edited and communicated Bayes's essay, published in *Philosophical Transactions* in 1763. The original experiment and later theorem should not be collapsed into modern Bayesian statistics. Pierre-Simon Laplace greatly expanded inverse-probability methods from the 1770s and in his 1812 *Théorie analytique des probabilités*. His famous perfect-intelligence image appears in the 1814 philosophical essay and expresses a deterministic setting in

which human probability tracks incomplete knowledge.

Bachelier, Einstein, Perrin and Wiener. Louis Bachelier's 1900 thesis used a continuous stochastic model for speculative prices. Albert Einstein's 1905 paper linked Brownian displacement to molecular activity and measurable diffusion. Jean Perrin's experiments and 1909 account, translated into English in 1910, supported the molecular interpretation. Norbert Wiener's 1923 "Differential-Space" supplied a rigorous measure on a space of continuous paths. The manuscript joins these as steps towards distributions over trajectories while keeping financial modelling, physical mechanism and experimental evidence distinct.

Hilbert and Kolmogorov. David Hilbert's sixth problem, presented in 1900, called for mathematical treatment of the axioms of physics and probability. Kolmogorov's 1933 monograph placed probability within measure theory using non-negativity, normalisation and countable additivity. The axioms govern formal consistency; they do not settle the interpretation of probability or validate any empirical model.

Markov and Eugene Onegin. In his 1913 study, Andrey Markov classified 20,000 letters from Pushkin's text as vowels or consonants. The modern English translation reports the sample and dependent sequence analysis. The counts of 8,638 vowels and 11,362 consonants are corroborated in Brian Hayes's historical reconstruction. The example demonstrates analysable dependence, not a modern language model in the current computational sense.

The modern Monte Carlo method. Work at Los Alamos associated with Stanislaw Ulam, John von Neumann, Nicholas Metropolis, Robert Richtmyer, Klara von Neumann and others developed computational statistical sampling for neutron transport and related problems. Metropolis's 1987 account reports Ulam's solitary recollection and the naming story. Laboratory and archival histories place the first computerised Monte Carlo calculations on ENIAC in April and May 1948, followed by later phases. Metropolis and Ulam published the general paper in 1949. Retrospective narratives differ in emphasis, so the body names a collaboration rather than

assigning sole invention.

What People Get Wrong and Use It

Single events and forecast quality. Calibration requires defined forecast cases, while proper scoring rules can compare individual probabilistic forecasts as well as systems. One outcome does not reveal a calibration rate, and selecting a forecast only after observing its outcome changes the question. The Brier score is one established method, though no single score captures every useful property.

Randomness and clustering. Independent random sequences contain runs. The probability of a pre-specified exact pattern differs from the probability that some striking pattern appears somewhere after many searches. The manuscript's examples keep that distinction visible and do not use "random" to mean evenly spaced.

Exposure and common causes. The complement formula converts per-opportunity risk into at-least-one-event risk only under the stated dependence structure. Under dependence, the joint model or scenario analysis must replace blind multiplication. Search by an adversary creates a different exposure process from accidental use.

Decision use. Expected value, quantiles, tail expectation, utility, reserves and reversibility answer different questions. No universal decision rule follows from probability alone. The practical lenses therefore require naming the event and horizon, rebuilding denominators, comparing likelihoods, carrying the distribution, and attacking the model before relying on simulation.

Causation. Conditional probability and Bayesian updating can compare causal models, but association alone does not identify intervention effects. The causal boundary is stated to preserve the deeper scope of *Statistics in a Hurry* and *Decision-Making in a Hurry*.

Terms support

The definitions follow standard probability usage. Density is distinguished from probability; expected value from a typical outcome; mutual exclusivity from independence; and a likelihood from a probability distribution over hypotheses. “Markov chain” is now included because dependence is load-bearing. Utility, tail expectation and more advanced constructions remain in the body without separate entries to preserve the thirty-term hierarchy.

Go Deeper verification

Publication details for the four recommendations were checked on 3 September 2026 against publisher or library records. Diaconis and Skyrms supply the broad conceptual tour; Hacking supplies a major history and philosophy; Feller supplies the mathematical classic; Gigerenzer supplies a practical treatment of conditional-risk communication. Their viewpoints differ, which is part of the recommendation design.

Bibliography

Primary and historical works

Bachelier, Louis. “Théorie de la spéculation.” *Annales scientifiques de l'École Normale Supérieure*, third series, 17 (1900): 21-86.

Bertrand, Joseph. *Calcul des probabilités*. Paris: Gauthier-Villars, 1889.

Bayes, Thomas, and Richard Price. “An Essay towards Solving a Problem in the Doctrine of Chances.” *Philosophical Transactions* 53 (1763): 370-418.

Bernoulli, Daniel. “Exposition of a New Theory on the Measurement of Risk.” Translated by Louise Sommer. *Econometrica* 22, no. 1 (1954): 23-36. Originally published 1738.

Bernoulli, Jacob. *The Art of Conjecturing, Together with Letter to a Friend on Sets in Court Tennis*. Translated by Edith Dudley Sylla. Baltimore: Johns Hopkins University Press, 2006. Originally published 1713.

Cardano, Girolamo. *The Book on Games of Chance (Liber de Ludo Aleae)*. Translated by Sydney Henry Gould. New York: Holt, Rinehart and Winston, 1961.

de Moivre, Abraham. *The Doctrine of Chances: Or, A Method of Calculating the Probabilities of Events in Play*. Third edition. London: A. Millar, 1756. First published 1718.

Einstein, Albert. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen." *Annalen der Physik* 322, no. 8 (1905): 549-560.

Graunt, John. *Natural and Political Observations Mentioned in a Following Index, and Made upon the Bills of Mortality*. London: John Martyn and James Allestry, 1662.

Halley, Edmond. "An Estimate of the Degrees of the Mortality of Mankind; Drawn from Curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to Ascertain the Price of Annuities upon Lives." *Philosophical Transactions* 17 (1693): 596-610.

Hilbert, David. "Mathematical Problems." Translated by Mary Winston Newson. *Bulletin of the American Mathematical Society* 8, no. 10 (1902): 437-479. Lecture delivered in 1900.

Huygens, Christiaan. *De ratiociniis in ludo aleae*. In Frans van Schooten, *Exercitationum mathematicarum libri quinque*, 517-534. Leiden: Johannes Elsevier, 1657.

Kolmogorov, A. N. *Foundations of the Theory of Probability*. Second English edition. Translated by Nathan Morrison. New York: Chelsea Publishing, 1956. Originally published 1933.

Laplace, Pierre-Simon. *Théorie analytique des probabilités*. Paris: Courcier, 1812.

Laplace, Pierre-Simon. *A Philosophical Essay on Probabilities*. Translated by Andrew I. Dale. New York: Springer, 1995. Based on the 1814 essay.

Markov, A. A. "An Example of Statistical Investigation of the Text *Eugene Onegin* Concerning the Connection of Samples in Chains." Translated by Gloria Custance and David Link. *Science in Context* 19, no. 4 (2006): 591-600. Original lecture delivered 1913.

Metropolis, Nicholas. "The Beginning of the Monte Carlo Method." *Los Alamos Science* 15, special issue (1987): 125-130.

Metropolis, Nicholas, and S. Ulam. "The Monte Carlo Method." *Journal of the American Statistical Association* 44, no. 247 (1949): 335-341.

Pacioli, Luca. *Summa de arithmetica, geometria, proportioni et proportionalita*. Venice: Paganino de Paganini, 1494.

Perrin, Jean. *Brownian Movement and Molecular Reality*. Translated by Frederick Soddy. London: Taylor and Francis, 1910.

Wiener, Norbert. "Differential-Space." *Journal of Mathematics and Physics* 2 (1923): 131-174.

Modern works

Bellhouse, David R. "Decoding Cardano's *Liber de Ludo Aleae*." *Historia Mathematica* 32, no. 2 (2005): 180-202.

Billingsley, Patrick. *Probability and Measure*. Third edition. New York: John Wiley & Sons, 1995.

Brier, Glenn W. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78, no. 1 (1950): 1-3.

Diaconis, Persi, and Frederick Mosteller. "Methods for Studying Coincidences." *Journal of the American Statistical Association* 84, no. 408 (1989): 853-861.

Diaconis, Persi, and Brian Skyrms. *Ten Great Ideas about Chance*. Princeton: Princeton University Press, 2018.

Edwards, A. W. F. "Pascal and the Problem of Points." *International Statistical Review* 50, no. 3 (1982): 259-266.

Feller, William. *An Introduction to Probability Theory and Its Applications*. Volume I. Third edition. New York: John Wiley & Sons, 1968.

Gigerenzer, Gerd. *Calculated Risks: How to Know When Numbers Deceive You*. New York: Simon & Schuster, 2002.

Hacking, Ian. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Second edition. Cambridge: Cambridge University Press, 2006.

Hájek, Alan. "Interpretations of Probability." *The Stanford Encyclopedia of Philosophy*. First published 2002; substantive revision 16 November 2023.

Hald, Anders. *A History of Probability and Statistics and Their Applications before 1750*. New York: John Wiley & Sons, 1990.

Hayes, Brian. "First Links in the Markov Chain." *American Scientist* 101, no. 2 (2013): 92-97.

Jaynes, E. T. "The Well-Posed Problem." *Foundations of Physics* 3 (1973): 477-492.

McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised edition. Princeton: Princeton University Press, 2015.

Robert, Christian P., and George Casella. *Monte Carlo Statistical Methods*. Second edition. New York: Springer, 2004.

Selvin, Steve. "A Problem in Probability." *The American Statistician* 29, no. 1 (1975): 67.

Selvin, Steve. "On the Monty Hall Problem." *The American Statistician* 29, no. 3 (1975): 134.

Seneta, Eugene. "A Tricentenary History of the Law of Large Numbers." *Bernoulli* 19, no. 4 (2013): 1088-1121.

Sood, Avneet, R. Arthur Forster, B. J. Archer, and R. C. Little. "Neutronics Calculation Advances at Los Alamos: Manhattan Project to Monte Carlo."

Nuclear Technology 207, supplement 1 (2021): S100-S133.