Jakob Bernoulli

On the Law of Large Numbers

Translated into English by Oscar Sheynin
Berlin 2005

Jacobi Bernoulli
Ars Conjectandi
Basileae, Impensis Thurnisiorum, Fratrum, 1713

Translation of Pars Quarta tradens *Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis*

[The Art of Conjecturing; Part Four showing *The Use and Application of the Previous Doctrine to Civil, Moral and Economic Affairs*]

ISBN 3-938417-14-5

Contents

Foreword

1. The *Art of Conjecturing* and Its Contents

Jakob Bernoulli (1654 – 1705) was a most eminent mathematician, mechanician and physicist. His *Ars Conjectandi* (1713) (AC) was published posthumously with a Foreword by his nephew, **Niklaus Bernoulli** (English translation: David (1962, pp. 133 – 135); French translation, J. Bernoulli (1987, pp. 11 – 12)). It is not amiss to add that N.Bernoulli (1709) published his dissertation on the application of the art of conjecturing to jurisprudence where he not only picked up some hints included in the manuscript of his late uncle, but borrowed whole passages both from it and even from the *Meditationes*, never meant for publication (Kohli 1975, p. 541).

The just mentioned *Meditationes* is Bernoulli's diary. It covers, approximately, the years 1684 – 1690 and is important first and foremost because it contains a fragmentary proof of the law of large numbers (LLN) to which Bernoulli indirectly referred at the end of Chapter 4 of Part 4 of the AC. Other points of interest in the *Meditationes* are that he (1975, p. 47) noted that the probability (in this case, statistical probability) of a visitation of a plague in a given year was equal to the ratio of the number of these visitations during a long period of time to the number of years in that period. Then, Bernoulli (p. 46, marginal note) wrote out the imprint of a review published in 1666 of **Graunt's** book (1662) which he possibly had not seen; he had not referred to it either in the *Meditationes* itself or in the AC. And, lastly, at about the same time Bernoulli (p. 43) considered the probability that an older man can outlive a young one (cf. Item 4 in Chapter 2, Part 4 of the AC). All this, even apart from the proof of the LLN, goes to show that already then he thought about applying statistical probability.

Part 1 of the AC is a reprint of **Huygens'** tract (1757) complete with vast and valuable commentaries. Nevertheless, this form testifies that Bernoulli was unable to complete his contribution. Also in Part 1 Bernoulli (pp. 22 – 28 of the German translation), while considering a game of dice, compiled a table which enabled him to calculate the coefficients of x^m in the development of $(x + x^2 + \ldots + x^6)^n$ for small values of n. Part 2 dealt with combinatorial analysis and it was there that the author introduced the *Bernoulli numbers.* Part 3 was devoted to application of the "previous" to drawing of lots and games of dice. Parts 1 and 3 contain interesting problems (the study of random sums for the uniform and the binomial distributions, a similar investigation of the sum of a random number of terms for a particular discrete distribution, a derivation of the distribution of the first order statistic for the discrete uniform distribution and the calculation of probabilities appearing in sampling without replacement). The author's analytical methods included combinatorial analysis and calculation of expectations of winning in each set of finite and infinite games and their subsequent summing.

Finally, Part 4 contained the LLN. There also we find a not quite formal "classical" definition of probability (a notion which he had not applied when formulating that law), a reasoning, in Chapter 2, on the aims of the art of conjecturing (determination, as precisely as possible, of probabilities for choosing the best solutions of problems, apparently in civil life) and elements of stochastic logic. Strangely enough, the title of Part 4 mentioned the completely lacking applications of the "previous doctrine" whereas his main theorem (the LLN) was not cited at all. This again testifies that Bernoulli had not completed his work. He did state, however (Chapter 4) that his LLN provided moral certainty which was sufficient for civil life and at the end of Chapter 2 he even maintained that judges must have firm instructions about what exactly constituted it.

Moral certainty had first appeared about 1400 (Franklin 2001, p. 69), but it was **Descartes** (1644, p. 323) who put it into circulation (above all apparently bearing in mind jurisprudence!). **Huygens** (Sheynin 1977, pp. 251 – 252) believed that proofs in physics were only probable and should be checked by appropriate corollaries and that common sense ought to determine the required degree of certainty of judgements in civil life. This latter statement seems much more reasonable than Bernoulli's rigid demand.

Bernoulli apparently considered the art of conjecturing as a mathematical discipline based on probability as a measure of certainty and on expectation and including (the not yet formally introduced) addition and multiplication theorems and crowned by the LLN.

2. The Art of Conjecturing, Part 4

2.1 Randomness and Necessity. Apparently not wishing to encroach upon theology, Bernoulli (beginning of Chapter 1) refused to discuss the notion of randomness. Then, again in the same chapter, he offered a subjective explanation of the "contingent" but actually corrected himself at the beginning of Chapter 4 where he explained randomness by the action of numerous complicated causes. Finally, the last lines of his book contain a statement to the effect that some kind of necessity was present even in random things (but left too little room for it). He referred to **Plato** who had taught that after a countless number of centuries everything returned to its initial state. Bernoulli likely thought about the archaic notion of the Great Year whose end will cause the end of the world with the planets and stars returning to their positions at the moment of creation. Without justification, he widened the boundaries of applicability of his law and his example was, furthermore, too complicated. It is noteworthy that **Kepler** (1596) believed that the end of the world was unlikely. In the first edition of this book his reasoning was difficult to understand but later he substantiated his conclusion by stating, in essence, like **Oresme** (1966, p. 247) did before him, that two [randomly chosen] numbers were "probably" incommensurable.

Bernoulli borrowed his example of finding a buried treasure from **Aristotle** (end of Chapter 1) but, unlike him, had not connected it with randomness. The later understanding of randomness began with **Maxwell** and especially **Poincaré**, who linked it with (among other interpretations) with the case in which slight causes (digging the earth somewhere near) would have led to considerable effects (the treasure remaining buried). Poincaré also sensibly reasoned on the interrelations between randomness and necessity. On the history of the notion of randomness see Sheynin (1991); new ideas took root late in the 20th century.

2.2. Stochastic Assumptions and Arguments. Bernoulli examined these in Chapters 2 and 3, but did not return to them anymore; he possibly thought of applying them in the unwritten pages of his book. The mathematical aspect of his considerations consisted in the use of the addition and the multiplication theorems for combining various arguments.

Unusual was the non-additivity of the deduced [probabilities] of the events under discussion. Here is one of his examples (Chapter 3, Item 7): "something" possesses 2/3 of certainty but its opposite has 3/4 of certainty; both possibilities are probable and their probabilities are as 8 : 9. Koopman (1940) resumed, in our time, the study of non-additive probabilities whose sources can be found in the medieval doctrine of probabilism that considered the opinion of each theologian as probable. Franklin (2001, p. 74) traced the origin of probabilism to the year 1577, or, in any case (p. 83), to 1611. Nevertheless, similar pronouncements on probabilities of opinion go back to **John of Salisbury** (the 12th century) and even to **Cicero** (Garber & Zabell 1979, p. 46).

I note a “general rule or axiom” concerning the application of arguments (pp. 234 and 236): out of two possibilities, the safer, the more reliable, etc should be chosen.

On the subject of this subsection see Shafer (1978) and Halperin (1988).

2.3. Arnauld and Leibniz. Antoine Arnauld (1612 – 1694) was an extremely well known religious figure and philosopher, the main author of the influential treatise **Arnauld & Nicole** (1662). In Chapter 4 Bernoulli praised Arnauld and approved his reasoning on using posterior knowledge and at the end of Chapter 3 Bernoulli borrowed Arnauld’s example (1662, pp. 328 – 329) of the criminal notary. Other points of interest are Arnauld’s confidence in moral certainty and his discussion of the application of arguments. It might be reasonably assumed that Arnauld was Bernoulli’s “non-mathematical” predecessor.

In 1703, Bernoulli informed **Leibniz** about the progress in his work (Kohli 1975, p. 509). He had been compiling it for many years with repeated interruptions caused by his “innate laziness” and worsening of health; the book still lacked its “most important part”, the application of the art of conjecturing to civil life; nevertheless, he, Bernoulli, had already shown his brother [Johann] the solution of a “difficult problem, special in its own way” that justified the applications of the art of conjecturing.

Most important both in that letter and in the following correspondence of 1703 – 1705 (Ibidem, pp. 510 – 512) was the subject of statistical probabilities. Leibniz never agreed that observations could secure moral certainty, but his arguments were hardly convincing. Thus, he in essence repeated the statement of Arnauld & Nicole (1662, pp. 304 and 317) that the finite (the mind; therefore, observations) could not always grasp the infinite (for example, God, but also, as Leibniz stated, any phenomenon depending on innumerable circumstances).

Leibniz’ views were possibly caused by his understanding of randomness as something “whose complete proof exceeds any human mind” (manuscript, 1686, p. 288). His heuristic statement does not contradict a modern approach to randomness founded on complexity and he was also right in the sense that statistical determinations cannot definitively corroborate a hypothesis.

In his letter of 3 Dec. 1703 Leibniz (Gini 1946, p. 405) also maintained that the allowance for all the circumstances was more important than subtle calculations, and **Bortkiewicz** (1923, p. 12) put on record **Keynes**’ (1921) favorable attitude towards this point of view and indicated the appropriate opinion of Mill (1843, p. 353), who had sharply contrasted the consideration of circumstances with “elaborate application” of probability and declared that the “neglect of this obvious reflection” made probability “the real opprobrium of mathematics”. Bortkiewicz agreed that mathematicians had been sometimes guilty of such neglect, which, however, had nothing to do with the calculus of probability. In his Chapter 4, Bernoulli touched on medical statistics and, for my part, I note that its progress is accompanied by the discovery of new circumstances so that stochastic calculations ought to be made repeatedly. Thus, in the mid-19th century, amputation of a limb made under the newly introduced anaesthesia sometimes led to death from bronchitis (Sheynin 1982, p. 262) and the benefits of that procedure had to be critically considered. *Circumstances* and *calculations* should not be contrasted.

Bernoulli paid due attention to Leibniz' criticism; more than a half of Chapter 4 of the AC in essence coincided with the respective passages from his letters to Leibniz (whom he did not mention by name).

In 1714, in a letter to one of his correspondents, Leibniz (Kohli 1975, p. 512) softened his doubts about the application of statistical probabilities and for some reason added that the late Jakob Bernoulli had "cultivated" the [theory of probability] in accordance with his, Leibniz' "exhortations".

On the correspondence between the two scholars see also Sylla (1998).

2.4. The Law of Large Numbers

2.4.1. The Prehistory. The LLN has its prehistory. It was thought, long before Bernoulli, that the number of successes in n "Bernoulli" trials with probability p was approximately equal to

$$\mu = np. \tag{1}$$

Cardano (Ore 1963, pp. 152 – 154 and 196), for example, applied this formula in calculations connected with games of dice. When compiling his mortality table, **Halley** (1694) assumed that "irregularities" in his data would have disappeared had he much more observations at his disposal. His idea can be interpreted as a statement on the increase in precision of formula (3) with n; it is likely, however, that these irregularities were occasioned by systematic corruptions. A second approach to the LLN took shape in astronomy not later than during **Kepler's** lifetime when the arithmetic mean became the universal estimator of the constant sought.

Similar but less justified statements concerning sums of magnitudes corrupted by random errors had also appeared. Thus, **Kepler** (Sheynin 1973, p. 120) remarked that the total weight of a large number of metal money of the same coinage did not depend on the inaccuracy in the weight of the separate coins. Then, **De Witt** (Sheynin 1977, p. 214) stated that the then existing custom of buying annuities upon many (n) young and apparently healthy lives secured profit "without hazard or risk". The expectation of a gain $\mathrm{E}x_i$ from each such transaction was obviously positive; if constant, the buyer could expect a total gain of $n\mathrm{E}x$. There also apparently existed a practice of an indirect participation of (petty?) punters in many games at once. At any rate (Sheynin 1977, p. 236), both **De Moivre** and **Montmort** mentioned in passing that some persons bet on the outcomes of games. The LLN has then been known, but not to such punters, and that practice could have existed from much earlier times.

2.4.2. Jakob Bernoulli. Before going on to prove his LLN, Bernoulli (Chapter 4) explained that the theoretical "number of cases" was often unknown, but what was impossible to obtain beforehand, might at least be determined afterwards, i.e., by numerous observations. In essence, Bernoulli proved a proposition that, beginning with **Poisson**, is being called the LLN. Let r and s be natural numbers, $t = r + s$, n, a large natural number, $\nu = nt$, the number of [independent] trials (De Moivre (1712) was the first to mention independence) in each of which the studied event occurs with [probability] r/t, μ – the number of the occurrences of the event (of the successes). Then Bernoulli proved without applying mathematical analysis that

$$P(|\frac{\mu}{\nu} - \frac{r}{t}| \le \frac{1}{t}) \ge 1 - \frac{1}{1+c} \tag{2}$$

and estimated the value of ν necessary for achieving a given $c > 0$. In a weaker form Bernoulli's finding meant that

$$\lim P\left(\left|\frac{\mu}{\nu} - \frac{r}{t}\right| < \varepsilon\right) = 1,\ \nu \to \infty \qquad (3)$$

where, as in formula (1), *r/t* was the theoretical, and μ/ν, the statistical probability.

Markov (1924, pp. 44 – 52) improved Bernoulli's estimate mainly by specifying his intermediate inequalities, and **Pearson** (1925), by applying the **Stirling** formula, achieved a practically complete coincidence of the Bernoulli result with the estimate that makes use of the normal distribution as the limiting case of the binomial law; Markov did not use that formula apparently because Bernoulli had not known it. In addition, Pearson (p. 202) considered Bernoulli's estimate of the necessary number of trials in formula (2) "crude" and leading to the ruin of those who would apply it. He also inadmissibly compared the Bernoulli law with the wrong Ptolemaic system of the world. The very fact described by formulas (2) and (3) was, however, extremely important for the development of probability and statistics, and, anyway, should we deny the importance of existence theorems? For modern descriptions of Bernoulli's LLN see Prokhorov (Bernoulli 1986) and Hald (1990, Chapter 16; 2003).

And so, the LLN established a correspondence between the two probabilities. Bernoulli (Chapter 4) had indeed attempted to ascertain whether or not the statistical probability had its "asymptote"– whether there existed such a degree of certainty, which observations, no matter how numerous, would never be able to reach. Or, in my own words, whether there existed such positive numbers β and δ < 1, that

$$\lim P\left(\left|\frac{\mu}{\nu} - \frac{r}{t}\right| < \beta\right) \le 1 - \delta,\ \nu \to \infty.$$

He answered his question in the negative: no, such numbers did not exist and he thus established, within the boundaries of stochastic knowledge, a relation between deductive and inductive methods and combined statistics with the art of conjecturing.

Throughout Part 4, Bernoulli considered the derivation of the statistical probability of an event given its theoretical probability and this most clearly emerges in the formulation of his *Main Proposition* in Chapter 5. However, both in the last lines of that chapter and in Chapter 4 he mentioned the inverse problem actually alleging that he had solved it as well. I return to this point in §2.4.3.

2.4.3. Remarks on Later Events. **De Moivre** (1765, p. 251) followed Bernoulli. Without any trace of hesitation, he claimed to have solved both the direct and the "converse" problems; he had expressed less clearly the same idea in 1738, in the previous edition of his book. **De Moivre's** mistake largely exonerates Bernoulli, so that **Keynes** (1921, p. 402) wrongfully stressed that the latter "proves the direct theorem only". It was **Bayes** who perceived that the two problems were different. He was the first to precisely determine the theoretical probability given the appropriate statistical data and for this reason

I (Sheynin 2003) suggested that **Bayes** had completed the construction of the first version of probability theory. This, however, does not diminish the great merit of Bernoulli in spite of the much more precise results of **De Moivre** (for one of the problems) and **Bayes**.

I do not discuss **Niklaus Bernoulli's** version of the LLN, which he described in one of his letters of 1713 to **Montmort** (1713, pp. 280 – 285); see Youshkevich (1986) and Hald (1990, §17.3; 2003). I myself (Sheynin 1970, p. 232; absent in the original publication of 1968) noted that **N.B.** was the first to have introduced, although indirectly, the normal distribution.

2.4.4. Alleged Difficulties in Application. Strangely enough, statisticians for a long time had not recognized the fundamental importance of the LLN. Haushofer (1872, pp. 107 – 108) declared that statistics, since it was based on induction, had no "intrinsic connections" with mathematics based on deduction [consequently, neither with probability]. A most noted German statistician, Knapp (1872, pp. 116 – 117), expressed a strange idea: the LLN was hardly useful since statisticians always made only one observation, as when counting the inhabitants of a city. And even later on, Maciejewski (1911, p. 96) introduced a "statistical law of large numbers" in place of the Bernoulli proposition that had allegedly impeded the development of statistics. His own law qualitatively asserted that statistical indicators exhibited ever lesser fluctuations as the number of observations increased.

All such statements definitely concerned the **Poisson** law as well (European statisticians then hardly knew about the **Chebyshev** form of the law) and Maciejewski's opinion likely represented the prevailing attitude of statisticians. Here, indeed, is what **Bortkiewicz** (1917, pp. 56 – 57) thought: the expression *law of large numbers* ought to be used only for denoting a "quite general" fact, unconnected with any definite stochastic pattern, of a higher or lower degree of stability of statistical indicators under constant or slightly changing conditions and given a large number of trials. Even **Romanovsky** (1961, p. 127) kept to a similar view and stressed the natural-science essence of the law and called it physical.

3. The Translations. Since 1713 the AC has appeared in a German translation whereas its Part 4 was translated into Russian and French (and, in an unsatisfactory way, into English); see the references. The German translation, especially insofar as mathematical reasoning is concerned, is rather far from the original; the Russian text also somewhat deviates from Bernoulli; finally, the French translation is perhaps almost faultless in this sense, but the translator made several mathematical mistakes. I do not read Latin and had to begin from the Russian text, but I invariably checked my work against the two other translations and the several English passages from the AC as provided by Shafer (1978) as well as against the original with the help of a Latin dictionary. I am really thankful to Claus Wittich (Geneve) who kindly went over my own text and translation and made valuable suggestions and corrections. I am confident that the final result is good enough but any remaining shortcomings and/or mistakes are my own.

A few words about **Markov** are in order. He initiated, and then edited the 1913 Russian translation mentioned above. The same year he put out the third, the jubilee edition, as he called it, of his treatise (see References) and supplied it with Bernoulli's portrait. Again in 1913, he initiated a special sitting of the Imperial [Petersburg] Academy of Sciences devoted to Bernoulli's work in probability and, along with two other mathematicians, delivered a report there, first published in 1914, reprinted in Bernoulli (1986) and available in an English translation (Ondar 1977/1981, pp. 158 –163). Later, in the posthumous edition of his treatise (1924), Markov improved Bernoulli's estimates (§2.4), and, perhaps as an indirect result of his study of the AC, inserted there many interesting historical comments.

The text of Part 4 of the *Art of Conjecturing* follows below.

Jakob Bernoulli. The Art of Conjecturing, Part 4 showing *The Use and Application of the Previous Doctrine to Civil, Moral and Economic Affairs*

Chapter 1. Some Preliminary Remarks about Certainty, Probability, Necessity and Fortuity of Things

Certainty of some thing is considered either *objectively* and in itself and means none other than its real existence at present or in the future; or *subjectively*, depending on us, and consists in the measure of our knowledge of this existence. Everything that exists or originates under the sun, – the past, the present, or the future, – always has in itself and objectively the highest extent of certainty. This is clear with regard to events of the present or the past; because, just by their existence or past existence, they cannot be non-existing or not having existed previously. Neither can you have doubts about [the events of] the future, which, likewise, on the strength of Divine foresight or predetermination, if not in accord with some inevitable necessity, cannot fail to occur in the future. Because, if that, which is destined to happen, is not certain to occur, it becomes impossible to understand how can the praise of the omniscience and omnipotence of the greatest Creator remain steadfast. But how can this certainty of the future be coordinated with fortuity or freedom [independence] of secondary causes? Let others argue about it; we, however, will not touch something alien to our aims.

Certainty of things, considered with respect to us, is not the same for all things, but varies diversely and occurs now greater, now lesser. Something, about which we know, either by revelation, intellect, perception, by experience, autopsia [direct observation; by one's own eyes] or otherwise, that we cannot in any way doubt its existence or realization in the future, has the complete and absolute certainty. To anything else our mind assigns a less perfect measure [of certainty], either higher or lower depending on whether there are more or less probabilities convincing us of its existence at present, in the past or the future.

As to *probability*, this is the degree of certainty, and it differs from the latter as a part from the whole. Namely, if the integral and absolute certainty, which we designate by letter α or by unity 1, will be thought to consist, for example, of five probabilities, as though of five parts, three of which favor the existence or realization of some event, with the other ones, however, being against it, we will say that this event has 3/5 α, or 3/5, of certainty. Therefore, the event having a greater part of certainty from among the other ones is called *more probable*, although actually, according to the usual word usage, we only call *probable* that, whose probability markedly exceeds a half of certainty. I say *markedly* because a thing, whose probability is roughly equal to a half of certainty, is called *doubtful* or indefinite. Thus, a thing having 1/5 of certainty is more probable than that which has 1/10, although actually neither one is probable.

Possible is that which has at least a low degree of certainty whereas the impossible has either no, or an infinitely small certainty. Thus, something is possible if it has 1/20 or 1/30 of certainty.

Morally certain is that whose probability is almost equal to complete certainty so that the difference is insensible. On the contrary, *morally impossible* is that which has only as much probability as the morally certain lacks for becoming totally certain. Thus, if morally certain is that which has 999/1000 of certainty, then something only having 1/1000 of certainty will be morally impossible.

Necessary is that, which cannot fail to exist at present, in the future or past, owing exactly to necessity, either *physical* (thus, fire will necessarily consume; a triangle will have three angles summing up to two right angles; a full moon, if in a node, will necessarily be accompanied by a [lunar] eclipse), – or *hypothetical*, according to which all that exists, or had existed, or is supposed to exist, cannot fail to exist (in this sense it is necessary that Petrus, about whom I know and accept that he is writing, is indeed writing), – or, finally, according to the *necessity* of a *condition* or *agreement* (thus, a gambler scoring a six with a die is necessarily reckoned the winner if the gamblers have agreed that winning is connected with throwing a six).

Contingent (both *free*, if it depends on the free will of a reasonable creature, and *fortuitous* and *casual*, if it depends on fortune or chance) is that which can either exist or not exist at present, in the past or future, – clearly because of remote rather than immediate forces. Indeed, neither does contingency always exclude necessity up to secondary causes. I shall explain this by illustrations. It is absolutely doubtless that, given a certain position of a die, [its] velocity and distance from the board at the moment when it leaves the thrower's hand, it cannot fall otherwise than it actually does. Just the same, under a certain present composition of the air, and given the masses, positions, motions, directions, and velocities of the winds, vapors and clouds, as well as the mechanical laws governing the interactions of all that, the weather tomorrow cannot be different from that which it will actually be. So these phenomena take place owing to their immediate causes with no lesser necessity than the phenomena of the eclipses follow from the movement of the heavenly bodies. And still, usually only the eclipses are ranked among necessary phenomena whereas the fall of a die and the future weather are thought to be contingent. The sole reason for this is that what is supposed to be known for determining future actions, and what indeed is such in nature, is not enough known. And, even had it been sufficiently known, geometrical and physical knowledge is inadequately developed for subjecting such phenomena to calculation in the same way as eclipses can be calculated beforehand and predicted by means of known astronomical principles. And, for the same reason, before astronomy achieved such perfection, the eclipses themselves had to be reckoned as future chance events to not a lesser extent than the two other [mentioned] phenomena.

It follows that what seems to be contingent to one person at a certain moment, will be thought necessary to someone else (or even to the same person) at another time after the [appropriate] causes become known. And so, contingency mainly depends on our knowledge since we do not see any contradiction with the non-existence of the event at present or in the future, although here and now, owing to an immediate but unknown to us cause, it is either necessarily realized, or ought to occur.

Not everything bringing us well-being or harm is called *happiness* or *misfortune* {*Fortuna prospera, un Bonheur*, ein Glück & *Fortuna adversa, un Malheur*, ein Unglück}, but only that which with a higher, or at least with the same probability could have not brought it. Therefore, happiness or misfortune are the greater, the lower was the probability of the well-being or harm that has actually occurred. Thus, exceptionally happy is the man who finds a buried treasure while digging the ground because this does not happen even once in a thousand cases. If twenty deserters, one of whom will be put to death by hanging as an example for the others, cast lots as to who remains living, those nineteen who drew the more favorable lot are not really called happy; but the twentieth who cast the horrible lot is most miserable. [In the same way,] your friend who came out unharmed from a battle in which [only] a small part of the combatants were killed should not be called happy, unless you will perhaps think it necessary to do so because of the special fortune of preserving life.

Chapter 2. On Arguments and Conjecture. On the Art of Conjecturing. On the Grounds for Conjecturing. Some General Pertinent Axioms

Regarding that which is certainly known and beyond doubt, we say that we *know* or *understand* [it]; concerning all the rest, – we only *conjecture* or *opine*.

To make *conjectures* about something is the same as to measure its probability. Therefore, the art of *conjecturing* or *stochastics* {*ars conjectandi* sive *stochastice*}[2.1] is defined as the art of measuring the probability of things as exactly as possible, to be able always to choose what will be found the best, the more satisfactory, serene and reasonable for our judgements and actions. This alone supports all the wisdom of the philosopher and the prudence of the politician.

Probabilities are estimated both by the *number* and the *weight of the arguments* which somehow prove or indicate that a certain thing is, was, or will be. As to the *weight*, I understand it to be the force of the proof.

Arguments themselves are either *intrinsic*, in every-day speech artificial, elicited in accordance with considerations of the cause, the effect, of the person, connection, indication or of other circumstances which seem to have some relation to the thing under proof; or *external* and not artificial, derived from people's authority and testimony. An example: Titius is found killed in the street. Maevius is charged with murder. The accusing arguments are: 1) He is known to have hated Titius (an argument from a *cause*, since this very hate could have incited to murder. 2) When questioned, he turned pale and answered timidly (this is an argument from the *effect* since it is possible that the pallor and fright were caused by his being conscious of the evil deed perpetrated). 3) Blood-stained cold steel is found in Maevius' house (this is an *indication*). 4) The same day that Titius was killed, Maevius had been walking the same road (this is *circumstance* of place and time). 5) Finally, Cajus maintains that the day before Titius was killed, he had quarrelled with Maevius (*this is a testimony*).

However, before getting down to our problem, – to indicating how should we apply these arguments for conjecturing so as to measure probabilities, – it is helpful to put forth some general rules or axioms which are dictated to any sensible man by usual common sense and which the more reasonable men always observe in everyday life.

1) *In such things in which it is possible to achieve complete certainty, there is no place for conjectures*. Futile would have been an astronomer, who, knowing that two or three [lunar] eclipses occur yearly, desires to forecast, on such grounds, whether or not there will be an eclipse during a full moon. Indeed, he could have found out the truth by reliable calculation. Just the same, if a thief says at his questioning that he sold the stolen thing to Sempronius, the judge who wants to conjecture about the probability of that statement by looking at the expression of the thief's face and listening to the tone of his voice, or by contemplating the quality of the stolen thing, or by some other circumstances, will act stupidly, because Sempronius, from whom everything can certainly and easily be elicited, is available.

2) *It is not sufficient to weigh one or another argument; it is necessary to investigate all such which can be brought to our knowledge and will seem to be suitable in some respect for proving the thing*. Suppose that three ships leave the harbor. After some time it is reported that one of them had suffered shipwreck and is lost. Conjectures are made: which of them? If only paying attention to the number of the ships, I shall conclude that each one of them could have met with the misfortune in an equal manner. But since I remember that one of them was comparatively old and decrepit, badly rigged with masts and sails, and steered by a young and inexperienced helmsman, I believe that, in all probability, it was this ship that got lost rather than one of the others.

3) *We ought to consider not only the arguments which prove a thing, but also all those which can lead to a contrary conclusion, so that, after duly discussing the former and the latter, it will become clear which of them have more weight*. It is asked, with respect to a friend very long absent from his fatherland, may we declare him dead? The following arguments favor an answer in the affirmative: During the entire twenty years, in spite of all efforts, we have been unable to find out anything about him; the lives of travellers are exposed to very many dangers from which those remaining at home are exempted; therefore, perhaps his life came to an end in the waves; perhaps he was killed en route or in battle; perhaps he died of an illness or from some [other] cause in a place where no one knew him. Then, has he been living, he would have reached an age which only a few attain even in their homeland; and he would have written even from the furthest shores of India because he knew that an inheritance was expected for him at home. And so on in the same vein.

Nevertheless, we should not rest content with these arguments but rather oppose them by the following supporting the contrary. He is known to have been thoughtless; wrote letters reluctantly; did not value friends. Perhaps Barbarians held him captive so that he was unable to write, or perhaps he did write sometimes from India, but the letters got lost either because of the carelessness of those carrying them, or during shipwrecks. And, to cap it all, many people are known to have returned unharmed after having been absent even longer[2.2].

4) *For judging about universalities remote and universal arguments are sufficient; however, for forming conjectures about particular things, we ought also to join to them more close and special arguments if only these are available*. Thus, if it is asked, in general, how much more probable is it for a

twenty-year-old youth to outlive an aged man of 60 rather than the other way round, we have nothing to take into consideration other than the distinction between the generations and ages. But if the question concerns two definite persons, the youth Petrus and the old man Paulus, we also ought to pay attention to their complexion, and to the care that each of them takes over his health. Because if Petrus is in poor health, indulges in passion, and lives intemperately, Paulus, although much older, may still hope, with every reason, to live longer.

5) *Under uncertain and dubious circumstances we ought to suspend our actions until more light is thrown. If, however, the necessity of action brooks no delay, we must always choose from among two possibilities that one which seems more suitable, safe, reasonable, or at least more probable*[2.3]*, even if none of them is actually such*. Thus, if a fire has broken out and you can only save yourself by jumping from the top of the roof or from some lower floor, it is better to choose the latter as being less dangerous, although neither alternative is quite safe or free from the danger of injury.

6) *That which is in some cases helpful and never harmful ought to be preferred to that which is never either helpful or harmful.* In our vernacular it is said *Hilfft es nicht, so schadt es nicht* [Even if it does not help, it does not harm]. This proposition follows from the previous [considerations], because that which can be helpful is more satisfactory, reliable and desirable than that which under the same conditions cannot [be helpful].

7) *Human actions should not be assigned a value according to their outcomes* because sometimes the most reckless actions are accompanied by the best success, whereas, on the contrary, the most reasonable [may] lead to the worst results. In agreement with this, the Poet says: “May success be wanting, I wish, for him who would judge facts by their outcomes” [Ovidius, Epistulae Heroidum II, “Phyllis Demophoonti”, line 85]. Thus, someone who intends to throw at once three sixes with three dice, should be considered reckless even if winning by chance. On the contrary, we [ought to] note the false judgement of the crowd which considers a man the more prominent, the more fortunate he is, and for which even a successful and fruitful crime is mostly a virtue. Once more Owen (*Epigr*[*ammatum*] *lib*[*er*] *sing*[*ularis*, 1607], §216)[2.4] gracefully says:

> *Although just now Ancus was believed to be a fool, it is argued that he is wise because the poorly conceived turned out successful* [*for him*]. *If something reasonably thought-out fails, even Cato will be judged a fool by the crowd.*

8) *In our judgements, we ought to beware of attributing to things more than is due to them, ought not to consider something which is only more probable than the other as absolutely certain, nor to impose the same opinion on others*. [This is] because the trust attributed to things ought to be in a proper proportion to the degree of certainty possessed by each thing, and be less in the same ratio as its probability itself is. In vernacular, this is expressed as *Man muss ein jedes in seinem Werth und Unwerth beruhen lassen* [Let each thing be determined by its value or worthlessness].

9) *However, since complete certitude can only seldom be attained, necessity and custom desire that that, which is only morally certain, be considered as absolutely certain*. Therefore, it would be helpful if the authorities determine

certain boundaries for moral certainty, – if, for example, it would be defined whether 99/100 of certainty be sufficient for resolving something, or whether 999/1000 be needed, so that a judge, unable to show preference to either side, will always have firm indications to conform with when pronouncing a sentence.

Anyone having knowledge of life can compile many more similar axioms but, lacking an appropriate occasion, we can hardly remember all of them.

Chapter 3. On Arguments of Different Kinds and on How Their Weights Are Estimated for Calculating the Probabilities of Things

He who considers various arguments by which our opinions and conjectures are formed will note a threefold distinction between them since some of them *necessarily exist and contingently provide evidence; others exist contingently and necessarily provide evidence; finally, the third ones both exist and provide evidence contingently.*

I explain these differences by examples. For a long time, my brother does not write me anything. I doubt whether to blame his laziness or his business pursuits, and fear that he may even have died. Here, there are threefold arguments for explaining the ceasing of the correspondence: *laziness, death, pursuits*. The first of these exists for sure (according to hypothetical necessity, since I know and accept that my brother is lazy), but proves true [provides evidence] only contingently because laziness possibly would not have hindered him from writing. The second one contingently exists (because my brother could still be alive), but proves true without question because a dead man cannot write. The third one both exists and provides evidence contingently because my brother can have business pursuits or not, and if he has them, they need not be such that prevent him from writing.

Another example. I suppose that, according to the conditions of a game, a gambler wins if he throws seven points with two dice, and I wish to guess his hope of winning. Here, the argument for winning is the throwing of seven points. It necessarily indicates the winning (owing, indeed, to the agreement between the gamblers), but it only exists contingently, because, in addition to the seven points, another number of them can occur.

Excepting this difference between the arguments, another distinction can also be noted since some of them are *pure*, the other ones, *mixed*. I call an argument *pure* if in some cases it proves a thing in such a manner that on other occasions it does not prove anything positively. A *mixed* argument, however, is such that in certain cases it thus proves a thing that on other occasions it proves the contrary in the same manner.

An example. Someone in a quarrelling crowd was stabbed with a sword; and, as trustworthy people who saw the incident from a distance testify, the perpetrator was dressed in a black cloak. If Gracchus was among those quarrelling together with three others, all of them in black tunics, this tunic will be an argument in favor of Gracchus having committed the murder. However, this argument will be mixed since in one case it proves his guilt, and, in three other cases, it demonstrates his innocence. Indeed, the murder was perpetrated either by him, or by one of the other three, with the latter instance being impossible without exonerating Gracchus. If, however, during the subsequent questioning Gracchus turned pale, the paleness of his face will be a pure argument because it demonstrates his guilt if occasioned by disturbed conscience. On the contrary, it would not prove his innocence had it been called forth by something else, since it is possible that he turned pale owing to another cause but still was himself the perpetrator of the murder [the murderer].

The above makes it clear that the force of proof peculiar to some argument depends on the multitude of cases in which it can exist or not exist, provide evidence or not, or even provide evidence to the opposite of the thing. Therefore, the degree of certainty, or the probability engendered by this argument, can be deduced by considering these cases in accordance with the doctrine given in Part 1 [of this book] in exactly the same way as the fate of gamblers in games of chance is usually investigated. To show it, let us assume that b is the number of cases in which some argument can accidentally exist, and c, in which it can fail to exist, with the number of both cases being $a = b + c$. Assume also that the number of cases in which the argument can contingently prove [the case] is β; and in which it does not prove it, or proves the opposite, γ, with the number of both cases being $\alpha = \beta + \gamma$. I suppose that all cases are equally possible and can take place with the same facility. Otherwise they ought to be moderated by assuming, instead of each easier occurring case, as many others as that case is easier to happen. For example, instead of a thrice easier case I will count three such cases that can occur as easily as the others.

1. And so, first, let *an argument exist contingently and provide evidence necessarily*. There will be b cases from among those just considered in which the argument can exist and therefore indicate a thing (or indicate 1), and c cases in which it can fail to exist and therefore to indicate. On the strength of Corollary 1 appended to Proposition 3 of Part 1 [of this book] the value [of the argument] will be

$$\frac{b \cdot 1 + c \cdot 0}{a} = \frac{b}{a}$$

so that such an argument proves (b/a) of a thing or [the same part] of its certainty.

2. Suppose now that *an argument necessarily exists but indicates contingently*. In accordance with the assumption, there will be β cases in which it can indicate the thing, and γ cases in which it does not indicate, or indicates the contrary. Here, the argument's force to prove the thing will therefore be

$$\frac{\beta \cdot 1 + \gamma \cdot 0}{\alpha} = \frac{\beta}{\alpha}.$$

Such an argument thus proves (β/α) of the thing's certainty; if, in addition, the argument is mixed, it proves (as is derived in the same way)

$$\frac{\gamma \cdot 1 + \beta \cdot 0}{\alpha} = \frac{\gamma}{\alpha}$$

of the contrary's certainty.

3. If some argument *both exists and indicates contingently*, I shall assume at first that it exists. Then, as just explained, it is estimated to prove (β/α) of the

thing, and, in addition, if the argument is a mixed one, to prove (γ/α) of the opposite. Since there are *b* cases in which it can exist; and fail to exist, and therefore to prove nothing, in *c* cases, it follows that this argument has value

$$\frac{b(\beta/\alpha)+c\cdot 0}{a}=\frac{b\beta}{a\alpha}.$$

If the argument is a mixed one, it has value

$$\frac{b(\gamma/\alpha)+c\cdot 0}{a}=\frac{b\gamma}{a\alpha}$$

of the opposite.

4. Then, if several arguments are collected for proving one and the same thing, and if we designate

The number of cases for argument No.	1	2	3	4	5	etc
The number of all of them	*a*	*d*	*g*	*p*	*s*	etc
The number of those proving	*b*	*e*	*h*	*q*	*t*	etc
And of those not proving or proving the opposite	*c*	*f*	*i*	*r*	*u*	etc

Then the force provided by the totality of all the arguments is estimated in the following way. First, suppose that all the arguments are *pure*. The weight of the first of them, taken separately, will be, as we saw, $(b/a) = (a - c)/a$; we should write (β/α) if the argument proves the thing only contingently, or ($b\beta/a\alpha$), if, in addition, it does not exist unquestionably. Now suppose that another argument is added to the first one, and that in *e*, or $(d - f)$ cases, it proves a thing, or proves 1, and in *f* cases proves nothing, with only the weight of the first argument, as found above, $[(a - c)/a]$, persisting. The weight of both arguments taken together will be

$$\frac{(d-f)\cdot 1+f[(a-c)/a]}{d}=1-\frac{cf}{ad} \text{ of the thing.}$$

Let us now add the third argument. There will be *h* or $(g - i)$ cases proving the thing, and *i* cases in which the third argument vanishes with only the force of the first two ones persisting. This force of proof is $[(ad - cf)/ad]$ from which it follows that the force of all three of them will be estimated as

$$\frac{(g-i)\cdot 1+i\,[(ad-cf)/ad]}{g}=1-\frac{cfi}{adg}.$$

And, when a larger number of arguments is available, we ought to go further in the same way. It is therefore clear that all the arguments taken together provide a probability differing from complete certainty or unity by a part of unity equal to the quotient of the product of the number of non-proving cases divided by the product of the number of all the cases for all the arguments.

5. Second, let all the arguments be *mixed*. Since the number of the proving cases for the first argument is b, for the second one, e, for the third argument, h, etc, and the number of cases proving the opposite is c, f, i, etc, it follows that, on the strength of only the first argument, the probability of the thing is to the probability of the opposite as b to c; according to the second one, as e to f; owing only to the force of the third argument, as h to i, etc. It is thus sufficiently clear that the total force of proof resulting from the totality of all the arguments is composed of the forces of all the separate arguments; or, that the probability of a thing is to the probability of the contrary as beh … to cfi … so that the absolute probability of the thing will be

$$\frac{beh}{beh+cfi}, \text{ and the absolute probability of the opposite, } \frac{cfi}{beh+cfi}.$$

6. Suppose that some arguments will again be *pure* (for example, the three first ones), and some *mixed* (for example, the two others). I consider at first only the pure arguments, which, according to Item 4, prove

$$\frac{adg-cfi}{adg}$$

of the thing's certitude. This fraction differs from 1 by cfi/adg, – as though, consequently, there be $(adg - cfi)$ cases in which these three arguments taken together prove the thing, or 1, and cfi cases in which they prove nothing and leave the proof to be solely accomplished by the mixed arguments. However, these latter two, on the strength of the proposition in Item 5, prove

$$\frac{qt}{qt+ru} \text{ of the thing, and } \frac{ru}{qt+ru} \text{ of the opposite.}$$

Therefore, the probability of the thing, following from all the arguments, is

$$\frac{(adg-cfi)\cdot 1+cfi\,[qt/(qt+ru)]}{adg}=\frac{adgqt+adgru-cfiru}{adgqt+adgru}=1-\frac{cfiru}{adg(qt+ru)},$$

which differs from complete certainty or 1 by the product of the fraction cf/adg (which, in accordance with Item 4, is the deficit of the probability [from unity] of the thing resulting from the pure arguments alone) by the fraction $ru/(qt + ru)$ that expresses the absolute probability of the opposite originating, as shown in Item 5, from the mixed arguments[3.1].

7. If, in addition to the arguments leading to the proof of a thing, there exist other pure arguments favoring the opposite, the arguments of both kinds, in accordance with the previous rules, ought to be weighed separately so as to derive the ratio in which the probability of the thing is to that of the opposite. And we ought to note here that, if the arguments pro and con are sufficiently strong, the absolute probabilities of each can appreciably exceed half of certainty; that is, either of the two opposite answers is probable, although one of them comparatively less probable than the other.

Thus, it might happen that something has 2/3 of certainty whereas its opposite has 3/4 so that each of these contraries will be probable although the first of them being less probable than the opposite; namely, their ratio will be as 2/3 to 3/4 or as 8 to 9.

I cannot conceal here that I foresee many obstacles in special applications of these rules that can often lead to shameful mistakes if caution is not observed when distinguishing between the arguments. Indeed, sometimes such arguments can seem to differ which actually compose one and the same argument, and to the contrary: differing arguments can be accepted as a single argument. Sometimes an argument includes such premises which absolutely refute the opposite, etc. As an explanation, I only adduce one or two illustrations. In the example above concerning Gracchus, I assume that the trustworthy people who saw those quarrelling also noted that the perpetrator was red-haired and that Gracchus together with two of the others were distinguished by hair of that color, but that no one of the latter was dressed in a black toga. In that case, if someone would have desired to compare the probabilities of Gracchus' guilt and innocence by the indications that Gracchus and three others were dressed in black, and also, that, again in addition to him, two others were notable for their red hair, and found that, according to Item 5, they are in a composite ratio of 1:3 and 1:2, or in the ratio of 1 to 6; and if he were to conclude that Gracchus is by far more likely to be innocent than to be the perpetrator of the murder, he would certainly have collated the matter in a most inept fashion. Actually, there are no two arguments here but only one and the same, resulting from two simultaneous circumstances, the color of the dress and of the hair. Since both these circumstances are only conjoined in the case of Gracchus, they certainly demonstrate that no one else excepting him could have been the perpetrator.

Another example. It becomes doubtful whether a written document is fraudulently antedated. An argument to the opposite could be that the document was signed by the hand of a notary public, i.e., by an official and sworn person, with regard to whom it is unlikely that he might have permitted himself any fraud. Indeed, he would have been unable to do so without greatly endangering his honor and well-being; in addition, even from among 50 notaries hardly one would have dared to commit such a vile action. The following arguments could be in favor of an answer in the affirmative: This notary is very ill-famed; and could have expected greatest benefits from the fraud; and especially that he had testified to something having no probability, as for example that someone had lent 10 000 gold coins to another person, whereas, according to everyone's estimation, all his property then barely amounted to 100.

Here, if considering separately the argument from the character of the signatory, the probability that the document is authentic may be valued as 49/50 of certainty. When, however, weighing the arguments favoring the opposite, it would be necessary to conclude that it is hardly possible that the document is not forged so that the fraud committed in the document is of course morally certain, that is, has 999/1000 of certainty. However, we should not conclude that the probabilities of authenticity and fraud are, in accordance with Item 7, in the ratio of 49/50 to 999/1000, or almost of equality. Because, if I believe that the notary is dishonorable, I am therefore assuming that he does not belong to the 49 honest notaries detesting deception but that he is indeed the fiftieth who has no scruples of fulfilling his duties faithlessly. This consideration completely destroys all the power of that argument, which in other cases could have been able to prove that a document is authentic.

Chapter 4. On a Two-Fold Method of Investigating the Number of Cases. What Ought To Be Thought about Something Established by Experience. A Special Problem Proposed in This Case, etc

It was shown in the previous chapter how, – given the number of cases in which arguments in favor of some thing can exist or fail to exist, can provide evidence or not, or even prove the opposite, – the force of what they prove, and the probabilities of things proportional to these forces, can be derived and estimated by calculation. We thus see that for correctly conjecturing about some thing, nothing else is required than both precisely calculating the number of cases and finding out how much more easily can some of them occur than the others. Here, however, we apparently meet with an obstacle since this only extremely seldom succeeds, and hardly ever anywhere excepting games of chance which their first inventors, desiring to make them fair, took pains to establish in such a way that the number of cases involving winning or losing were determined with certainty and known and the cases themselves occurred with the same facility.

However, for most of other matters, depending either on the production of nature or the free will of people, this does not take place at all. Thus, for example, the number of cases is known in [a game of] dice. For each die there are manifestly as many cases as faces, and all of them are equally inclined [to turn up], since, owing to the similitude [congruence] of the faces and the uniform weight [density] of the die, there is no reason for one of them to turn up more easily than another[4.1]. This would have happened if the forms of the faces were dissimilar or if one part of the die consisted of a heavier substance than the other one. In the same way, the number of cases for drawing a white or a black ticket from an urn is known, and known [also] is that [the drawings of] all of them are equally possible. Indeed, the number of tickets of both these kinds is evidently determined and known, and no reason is seen for one of them to appear more easily than any other.

But, who from among the mortals will be able to determine, for example, the number of diseases, that is, the same number of cases which at each age invade the innumerable parts of the human body and can bring about our death; and how much easier one disease (for example, the plague) can kill a man than another one (for example, rabies; or, the rabies than fever), so that we would be able to conjecture about the future state of life or death? And who will count the innumerable cases of changes to which the air is subjected each day so as to form a conjecture about its state in a month, to say nothing about a year? Again, who knows the nature of the human mind or the admirable fabric of our body shrewdly enough for daring to determine the cases in which one or another participant can gain victory or be ruined in games completely or partly depending on acumen or agility of body?

Since this and the like depends on absolutely hidden causes, and, in addition, owing to the innumerable variety of their combinations always escapes our diligence, it would be an obvious folly to wish to find something out in this manner. Here, however, another way for attaining the desired is really opening for us. And, what we are not given to derive a priori, we at least can obtain *a posteriori*, that is, can extract it from a repeated observation of the results of similar examples. Because it should be assumed that each phenomenon can occur and not occur in the same number of cases in which, under similar circumstances, it was previously observed to happen and not to happen. Actually, if, for example, it was formerly noted that, from among the observed three hundred men of the same age and complexion as Titius now is and has, two hundred died after ten years with the others still remaining alive, we may conclude with sufficient confidence that Titius also has twice as many cases for paying his debt to nature during the next ten years than for crossing this border. Again, if someone will consider the atmosphere for many previous years and note how many times it was fine or rainy; or, will be very often present at a game of two participants and observe how many times either was the winner, he will thus discover the ratio of the number of cases in which the same event will probably happen or not also in the future under circumstances similar to those previously existing.

This empirical method of determining the number of cases by experiment is not new or unusual. Because the celebrated author of *L'art de penser*, a man of great intellect and acumen[4.2], prescribes the like in Chapter 12 and in the next ones of the last part [of that book], and the same is also constantly observed in everyday practice. Then, neither can anyone fail to note also that it is not enough to take one or another observation for such a reasoning about an event, but that a large number of them are needed. Because, even the most stupid person, all by himself and without any preliminary instruction, being guided by some natural instinct (which is extremely miraculous) feels sure that the more such observations are taken into account, the less is the danger of straying from the goal.

Although this is known by nature to everyone, its proof, derived from scientific principles, is not at all usual and we ought therefore to expound it here. However, I would have estimated it as a small merit had I only proved that of which no one is ignorant. Namely, it remains to investigate something that no one had perhaps until now run across even in his thoughts. It certainly remains to inquire whether, when the number of observations thus increases, the probability of attaining the real ratio between the number of cases, in which some event can occur or not, continually augments so that it finally exceeds any given degree of certitude. Or [to the contrary], the problem has, so to say, an asymptote; i.e., that there exists such a degree of certainty which can never be exceeded no matter how the observations be multiplied, so that, for example, it is never possible to obtain more than a half, or than 2/3, or 3/4, of certainty in deriving the real ratio of cases.

To make clear my desire by illustration, I suppose that without your knowledge three thousand white pebbles[4.3] and two thousand black ones are hidden in an urn, and that, to determine [the ratio of] their numbers by experiment, you draw one pebble after another (but each time returning the drawn pebble before extracting the next one so that their number in the urn will not decrease), and note how many times is a white pebble drawn, and how many times a black one. It is required to know whether you are able to do it so many times that it will become ten, a hundred, a thousand, etc., times more probable (i.e., become at last morally certain) that the number of the white and the black pebbles which you extracted will be in the same ratio, of 3 to 2, as the number of pebbles themselves, or cases, than in any other different ratio. To tell the truth, if this failed to happen, it would be necessary to question our attempt at experimentally determining the number of cases.

If, however, this is attained and we thus finally obtain moral certainty (in the next chapter I shall show that this is indeed so), then we determine the number of cases *a posteriori* almost as though it was known to us a priori. In social life, where the morally certain, according to Proposition 9 of Chapter 2, is assumed as quite certain, this is undoubtedly quite sufficient for scientifically directing our conjectures about any contingent thing in a no lesser way than in games of chance. Because, if we replace an urn for example by air or by a human body, which contain in themselves sources of various changes or diseases just as the urn contains pebbles, we will be able to determine by observation in exactly the same way how much easier can one or another event occur in these things.

To avoid false understanding, it ought to be noted that the ratio between the numbers of cases which we desire to determine experimentally is accepted not as precise and strict (because this point of view would have led to a contrary result and the probability of determining the real ratio would have been the lower the more observations we would have taken)[4.4], but that this ratio be accepted with a certain latitude, that is, contained between two limits [boundaries] which could be taken as close as you like. Indeed, if in the example just provided concerning pebbles, we will assume two ratios, 301/200 and 299/200, or 3001/2000 and 2999/2000, etc, one of which is very near but greater, and the other one very near but smaller than 3/2, it will be shown that, setting any probability, it can be made more probable that the ratio derived from many observations will be contained within these limits of 3/2 rather than outside.

This, then, is the problem that I decided to make here public after having known its solution for twenty years. Its novelty and the greatest utility joined with an equal difficulty can attach more weight and value to all the other chapters of this doctrine [of the ars conjectandi]. However, before exposing its solution I shall defend myself in a few words from the objections to these propositions levelled by some scholars.

1. First, it was objected that the ratio of pebbles is one thing, whereas the ratio of diseases or changes in the air is something else. The number of the former is definite but the number of the latter is indefinite and vague. I answer this by saying that they both, in comparison to our knowledge, are equally indefinite and vague. However, we can imagine anything, that is such in itself and in accordance with its nature, not better than a thing created and at the same time not created by the Author of nature because everything done by God is determined thereby.

2. Second, it is objected that the number of pebbles is finite and that of diseases etc. is infinite. *Answer*. Rather immense than infinite. But let us assume that it is indeed infinite. Even between two infinities a definite ratio is known to be possible and to be expressed by finite numbers either precisely or at least with any desired approximation. Thus, the ratio of each circumference to [its] diameter is definite. [True,] it is not precisely expressed otherwise than by an infinitely continued Ludolphus' cyclic number. However, Archimedes, Metius and Ludolphus himself[4.5] contained that ratio within limits [boundaries] sufficiently close to each other for practice. Therefore, nothing hinders a ratio of two infinities approximately expressed by finite numbers to be determined by a finite number of experiments either.

3. Third, it is said that the number of diseases does not remain constant but that new diseases occur every day. *Answer*. We are unable to deny that

diseases can multiply in the course of time; and he who desires to conclude from present-day observations about the times of our antediluvian forefathers will undoubtedly deviate enormously from the truth. But nothing follows from this except that sometimes we ought to resume observations just as it would be necessary to resume observations with the pebbles if it is assumed that their number in the urn is variable.

Chapter 5. Solution of the Previous Problem

To explicate the long demonstration as briefly and clearly as possible, I will attempt to reduce everything to abstract mathematics, eliciting from it the following lemmas after which all the rest will only consist in their mere application.

***Lemma* 1.** Suppose that a series of any quantity of numbers 0, 1, 2, 3, 4, etc, follow, beginning with zero, in the natural order and let the extreme and maximal of them be $r + s$, some intermediate, r, and the two nearest to it on either side, $r + 1$ and $r - 1$. If this series be continued until its extreme term becomes equal to some multiple of the number $r + s$, that is, until it is equal to $nr + ns$, the intermediate number r and those neighboring it, $r + 1$ and $r - 1$, will be augmented in the same ratio, so that nr, $nr + n$ and $nr - n$ will appear instead, and the series itself

$$0, 1, 2, 3, 4, \ldots, r-1, r, r+1, \ldots, r+s$$

will change becoming

$$0, 1, 2, 3, 4, \ldots, nr-n, \ldots, nr, \ldots, nr+n, \ldots, nr+ns.$$

With an increasing n both the number of the terms situated between the intermediate nr and one of the limiting terms, $nr + n$ or $nr - n$, and the number of those terms which extend from these limits to the extreme terms $nr + ns$ or 0 will thus increase. However (no matter how large will n be assumed), the number of terms following after the larger limit $nr + n$ will never be more than $s - 1$ times greater than, and the number of terms preceding the lesser limit $nr - n$ will never be more than $r - 1$ times greater than the number of them situated between the intermediate nr and one of the limits, $nr + n$ or $nr - n$. Because, after subtraction, it is clear that between the greater limit and the extreme term $nr + ns$ there are $ns - n$ intermediate terms, and between the lesser limit and the other extreme term 0 there are $nr - n$ intermediate terms, and n terms between the intermediate and each of the limits. However, $(ns - n) : n = (s - 1):1$ and $(nr - n):n = (r - 1):1$. It therefore follows, etc.

***Lemma* 2.** A binomial $r + s$ raised to any integral power is expressed by terms whose number exceeds by 1 the number of unities in the exponent. Because a square [of a binomial] consists of three terms, a cube has 4, a fourth power has 5 terms, etc, as is known.

***Lemma* 3.** For any power of this binomial (at least for an exponent equal to the binomial $r + s = t$, or to its multiple, for example, to $nr + ns = nt$), a certain term M will be maximal if the number of terms preceding and following it are in the ratio of s to r; or, which is the same, if the exponents of letters r and s in this term are in the ratio of the magnitudes r and s themselves. The term nearer to it from either side is greater than the more distant term on the same side; however, the same term *M* is in a lesser ratio to the nearer term than the nearer term to the more distant one if the numbers of intermediate terms are the same.

Dem[*onstration*]. 1. Geometers know well enough that the binomial $r + s$ raised to the power nt, that is, $(r + s)^{nt}$, is expressed by such a series:

$$r^{nt} + \frac{nt}{1} r^{nt-1} s + \frac{nt(nt-1)}{1\cdot 2} r^{nt-2} s^2 + \frac{nt(nt-1)(nt-2)}{1\cdot 2\cdot 3} r^{nt-3} s^3 + \text{etc until}$$

$$\frac{nt}{1} rs^{nt-1} + s^{nt}.$$

Here, the powers of *r* gradually decrease and those of *s* increase, whereas the coefficients of the second and the last but one terms become $nt/1$; of the third counting from the beginning and the end, $[nt(nt - 1)/1\cdot 2]$; of the fourth from the beginning and the end, $[nt(nt - 1)(nt - 2)/ 1\cdot 2\cdot 3]$, etc.

Since the number of all the terms excepting *M* is, according to Lemma 2, $nt = nr + ns$, and, as assumed, the numbers of terms preceding and following *M* are as *s* to *r*, these numbers are ns and nr respectively. Therefore, in accordance with the law of the [formation of the] series, the term *M* will be

$$\frac{nt(nt-1)(nt-2)...(nr+1)}{1\cdot 2\cdot 3\cdot 4\ldots ns} r^{nr} s^{ns} \text{ or } \frac{nt(nt-1)(nt-2)...(ns+1)}{1\cdot 2\cdot 3\cdot 4\ldots nr} r^{nr} s^{ns}$$

and, in the same way, the terms nearest to it on *the left* and *the right* are

$$\frac{nt(nt-1)(nt-2)...(nr+2)}{1\cdot 2\cdot 3\cdot 4\ldots (ns-1)} r^{nr+1} s^{ns-1} \text{ and } \frac{nt(nt-1)(nt-2)...(ns+2)}{1\cdot 2\cdot 3\cdot 4\ldots (nr-1)} r^{nr-1} s^{ns+1}$$

and in the same way the next ones on *the left* and *the right* are

$$\frac{nt(nt-1)(nt-2)...(nr+3)}{1\cdot 2\cdot 3\cdot 4\ldots (ns-2)} r^{nr+2} s^{ns-2} \text{ and } \frac{nt(nt-1)(nt-2)...(ns+3)}{1\cdot 2\cdot 3\cdot 4\ldots (nr-2)} r^{nr-2} s^{ns+2}.$$

After a preliminary suitable cancellation of common multipliers from both the coefficients and the powers themselves, it becomes clear that the term *M* is to its nearest on the left as $(nr + 1)s$ to nrs; this latter to the next one, as $(nr + 2)s$ to $(ns - 1)r$, etc, and also that the term *M* is to its nearest on the right as $(ns + 1)r$ to nsr, this latter to the next one, as $(ns + 2)r$ to $(nr - 1)s$, etc. But

$(nr + 1)s > nrs$, and $(nr + 2)s > nsr - r$, etc.

Also,

$(ns + 1)r > nsr$ and $(ns + 2)r > nrs - s$, etc.

It follows that the term *M* is greater than either of the nearest terms on either side which [in turn] are greater than the more remote terms on the same side, etc. QED.

2. The ratio $(nr + 1)/ns$, as is clear, is less than the ratio $(nr + 2)/(ns - 1)$. Therefore, after multiplying [them] by one and the same ratio s/r, the ratio

$$\frac{(nr+1)s}{nsr} < \frac{(nr+2)s}{(ns-1)r}.$$

Just the same, it is evident that the ratio

$$\frac{ns+1}{nr} < \frac{ns+2}{nr-1}.$$

Consequently, after multiplying [this inequality] by one and the same ratio *r/s*, also

$$\frac{(ns+1)r}{nrs} < \frac{(ns+2)r}{(nr-1)s}.$$

But the ratio

$$\frac{(nr+1)s}{nsr}$$

is equal to the ratio of the term *M* to its nearest term on the left and the ratio[5.1]

$$\frac{(nr+2)s}{(ns-1)r}$$

is the same as that has to the next one. And the ratio

$$\frac{(ns+1)r}{nrs}$$

is that of the term *M* to its nearest term on the right, and

$$\frac{(ns+2)r}{(nr-1)s}$$

is the ratio of that term to the next one. What was just shown may in the same way be also applied to all the other terms.

Therefore, the maximal term *M* is in a lesser ratio to the nearer term on either side than (if the intervals between the terms are the same) the nearer term is to the more distant one on the same side. QED.

***Lemma* 4.** The number *n* in a binomial raised to the power *nt* can be taken so great that the ratio of the maximal term *M* to [any of the] two others, *L* and Λ, distant from it by *n* terms on the left and on the right [respectively], would be greater than any given ratio.

Dem[*onstration*]. Since in the previous Lemma the maximal term *M* was found to be equal to

$$\frac{nt(nt-1)(nt-2)...(nr+1)}{1\cdot 2\cdot 3\cdot 4\ldots ns} r^{nr} s^{ns} \text{ or } \frac{nt(nt-1)(nt-2)...(ns+1)}{1\cdot 2\cdot 3\cdot 4\ldots nr} r^{nr} s^{ns},$$

the terms on the left and on the right, L and Λ, in accordance with the law of the [formation of the] series (adding n to the last multiplier in the numerators of the coefficients, and subtracting n from the last multiplier in their denominators, adding the same n to the power of one of the letters r and s, and subtracting it from the power of the other letter), will be

$$\frac{nt(nt-1)(nt-2)...(nr+n+1)}{1\cdot 2\cdot 3\cdot 4\,...\,(ns-n)}r^{nr+n}s^{ns-n} \text{ and}$$

$$\frac{nt(nt-1)(nt-2)...(ns+n+1)}{1\cdot 2\cdot 3\cdot 4\,...\,(nr-n)}r^{nr-n}s^{ns+n},$$

so that, after a suitable cancellation of common multipliers,

$$\frac{M}{L}=\frac{(nr+n)(nr+n-1)(nr+n-2)...(nr+1)s^n}{(ns-n+1)(ns-n+2)(ns-n+3)\,...\,nsr^n},$$

$$\frac{M}{\Lambda}=\frac{(ns+n)(ns+n-1)(ns+n-2)...(ns+1)r^n}{(nr-n+1)(nr-n+2)(nr-n+3)...nrs^n},$$

or (with rn and sn being equally distributed among the [other] multipliers, equal in number)

$$\frac{M}{L}=\frac{(nrs+ns)(nrs+ns-s)(nrs+ns-2s)...(nrs+s)}{(nrs-nr+r)(nrs-nr+2r)(nrs-nr+3r)...nrs},$$

$$\frac{M}{\Lambda}=\frac{(nrs+nr)(nrs+nr-r)(nrs+nr-2r)...(nrs+r)}{(nrs-ns+s)(nrs-ns+2s)(nrs-ns+3s)...nrs}.$$

However, when n is assumed infinite, these ratios will [also] be infinitely large, because then the numbers 1, 2, 3 etc will vanish as compared with n, and the numbers themselves $nr \pm n \mp 1$, $nr \pm n \mp 2$, $nr \pm n \mp 3$, etc, and $ns \pm n \mp 1$, $ns \pm n \mp 2$, $ns \pm n \mp 3$, etc will have the same value as $nr \pm n$ and $ns \pm n$ [respectively], so that, after dividing [both parts of both last fractions] by n,

$$\frac{M}{L}=\frac{(rs+s)(rs+s)(rs+s)...rs}{(rs-r)(rs-r)(rs-r)...rs},\ \frac{M}{\Lambda}=\frac{(rs+r)(rs+r)(rs+r)...rs}{(rs-s)(rs-s)(rs-s)...rs}.$$

It is clear that these ratios are composed of as many ratios [$(rs+s)/(rs-r)$] or [$(rs+r)/(rs-s)$] as there are multipliers whose number is n, that is, infinite because the difference between the first multipliers $nr+n$ or $ns+n$, and the last ones, $nr+1$ or $ns+1$, is $n-1$. These ratios [M/L and M/Λ] will therefore be equal to [$(rs+s)/(rs-r)$] or [$(rs+r)/(rs-s)$] raised to an infinite power and therefore simply infinite. If you doubt this conclusion, imagine an infinity [of ratios] in a continued proportion with their ratio being as $rs+s$ to $rs-r$ or $rs+r$ to $rs-s$. The first ratio will be to the third as the square; to the fourth, as a cube; to the fifth, as the fourth [power], etc. Finally, the first ratio will be to the the last one as infinite powers of the ratio $[(rs + s)/(rs - r)]$ or $[(rs + r)/(rs - s)]$. It is known, however, that the ratio of the first [ratio] to the last one is infinitely large since the last one = 0 (*see Coroll. to Prop*[*osition*] 6 *of our* [*Tractatus de*] *Seriebus Infinitis* [etc][5.2]). It is therefore clear that infinite powers of the ratio $[(rs + s)/(rs - r)]$ or $[(rs + r)/(rs - s)]$ are infinite. It is thus shown that the ratio of the maximal term M to [any of the] two others, L and Λ, exceeds any assigned ratio. QED.

***Lemma* 5.** Assuming the same as above, it is possible to imagine such a large number n, that the sum of all the terms from the intermediate and maximal M to both the [to any of the] terms L and Λ inclusive, is to the sum of all the other terms exterior to the limits L and Λ, in a ratio greater than any given ratio.

Dem[*onstration*]. Let the terms between the maximal M and the limiting term L on the left be designated: the second one from the maximal[5.3], F, the third one, G, the fourth one, H, etc; and the second one beyond L, P, the third one, Q, the fourth one, R, etc. Since according to the second part of Lemma 3 the ratios

$$M/F < L/P,\ F/G < P/Q,\ G/H < Q/R, \text{ etc,}$$

we will also have

$$M/L < F/P < G/Q < H/R, \text{ etc.}$$

Since, according to Lemma 4, the ratio M/L for an infinitely large n is infinite, the ratios F/P, G/Q, H/R, … will all the more be infinite and therefore the ratio

$$\frac{F+G+H+...}{P+Q+R+...}$$

is also infinite; that is, the sum of the terms between the maximal term M and the limit L is infinitely greater than the sum of the same number of terms beyond and nearest to L. And since according to Lemma 1 the number of all the terms outside L is not more than $s - 1$ times (i.e., not more than a finite number of times) greater than the number of terms between this limit and the maximal term M, and the terms themselves, in accordance with the first part of Lemma 3, become the smaller the further they are from the limit, the sum of all the terms between M and L (even without considering M) will be infinitely greater than the sum of all the terms beyond L. In a similar way it is shown that the sum of all the terms between M and Λ is infinitely greater than the sum of all the terms beyond Λ (whose number, according to Lemma 1, is not more than $r - 1$ times greater than the number of the former). Therefore, finally, the sum of all the terms situated between the limits L and Λ (the maximal term may be excluded) will be infinitely greater than the sum of all the terms beyond these limits. Consequently, this statement persists all the more if the maximal term is included [in the first sum], QED.

***Explanatory Comment*.** Those, who are not acquainted with inquiries involving infinity may object to Lemmas 4 and 5, That, – although, if n is infinite, the multiples of the magnitudes expressing the ratios M/L and M/Λ, that is, $nr \pm n \mp 1$, $nr \pm n \mp 2$, $nr \pm n \mp 3$, etc, and $ns \pm n \mp 1$, $ns \pm n \mp 2$, ns

$\pm n \mp 3$, etc, have the same value as $nr \pm n$ and $ns \pm n$ since numbers 1, 2, 3, … vanish with respect to each multiplier, – it can still happen that, taken together and multiplied one by another, they increase to infinity (because the number of multipliers is infinite) and will infinitely decrease, that is, make finite, the infinite powers of the ratios [$(rs + s)/(rs - r)$] or [$(rs + r)/(rs - s)$]. I cannot obviate these scruples better than by showing now a method of actually deriving a finite number n, or a finite power of a binomial, for which the sum of the terms between the limits L and Λ has a larger ratio to the sum of the terms beyond them than any no matter how great given ratio, which I designate by letter c. Once this is shown, the objection will necessarily fall away.

To this end, I choose some ratio [greater than unity] less, however, than the ratio [$(rs + s)/(rs - r)$] (for the terms on the left), – for example, the ratio [$(rs + s)/rs$] or $(r + 1)/r$, – and multiply it by itself so many times (m times) that the product becomes equal or exceeds the ratio of $c(s - 1)$ to 1; that is, until

$$[(r + 1)^m / r^m] \geq c(s - 1).$$

When will this happen can be advantageously investigated by means of logarithms. Because, taking logarithms, we obtain

$$m \operatorname{Log}(r + 1) - m \operatorname{Log} r \geq \operatorname{Log} [c(s - 1)]$$

and, after dividing, we find at once that

$$m \geq \frac{\operatorname{Log} [c(s-1)]}{\operatorname{Log} (r+1) - \operatorname{Log} r}.$$

Having found this, I continue in the following way. With regard to a series of fractions or multipliers

$$\frac{nrs + ns}{nrs - nr + r}, \frac{nrs + ns - s}{nrs - nr + 2r}, \frac{nrs + ns - 2s}{nrs - nr + 3r}, \ldots, \frac{nrs + s}{nrs},$$

from which, according to Lemma 4, the ratio M/L is obtained by multiplying them one by another, it may be remarked that, although the separate fractions are less than the fraction [$(rs + s)/(rs - r)$], they approach it the nearer the larger is the assumed n. Therefore, one of them will sooner or later become equal to the ratio [$(rs + s)/rs$] = [$(r + 1)/r$] itself. It should be therefore found out how great n ought to be taken for the fraction whose ordinal number is m to become equal to [$(r + 1)/r$] itself. But (as it is seen from the law of the formation of the series) the fraction of ordinal number m is

$$\frac{nrs + ns - ms + s}{nrs - nr + mr}.$$

Equating it to [$(r + 1)/r$], we obtain

$$n = m + \frac{ms - s}{r+1} \text{ so that } nt = mt + \frac{mst - st}{r+1}.$$

I maintain that if this is the power to which the binomial $(r + s)$ is raised, the maximal term M will be more than $c(s - 1)$ times greater than the limit L. Indeed, for the assumed value of n the fraction of ordinal number m will be equal to $[(r + 1)/r]$, and the fraction $[(r + 1)/r]$, being multiplied by itself m times, that is [the fraction] $[(r + 1)^m/r^m]$, is (as constructed) equal or greater than $c(s - 1)$. Therefore, this fraction [of ordinal number m] multiplied by all the previous fractions will all the more exceed $c(s - 1)$ since all these are greater than $[(r + 1)/r]$.

Consequently, the product, being multiplied by all the following [fractions], will all the more exceed $c(s - 1)$ because each of these is at least greater than unity. But the product of all the fractions expresses the ratio of the term M to term L and it is therefore absolutely clear that the term M exceeds the limit L over $c(s - 1)$ times.

But, as was shown,

$$M/L < F/P < G/Q < H/R, \text{ etc.}$$

It follows that the second term after the maximal term M exceeds the second term after the limit L more than $c(s - 1)$ times, that the third term [after M] still more exceeds the third term [after L], etc. Therefore, finally, the sum of all the terms between the maximal M and the limit L will exceed the sum of the same number of maximal terms situated beyond this limit more than $c(s - 1)$ times, and more than c times the same sum taken $(s - 1)$ times. Consequently, it is still more evident that it exceeds more than c times the sum of all the terms situated beyond the limit L whose number is not more than $s - 1$ times greater [than the number of terms between M and L].

I proceed in the same way with regard to the terms on the right. I take the ratio

$$\frac{s+1}{s} < \frac{rs+r}{rs-s},$$

assume that

$$\frac{(s+1)^m}{s^m} \geq c(r-1)$$

and determine

$$m \geq \frac{\text{Log}\,[c(r-1)]}{\text{Log}\,(s+1) - \text{Log}\,s}.$$

Then, from among the series of fractions

$$\frac{nrs+nr}{nrs-ns+s}, \frac{nrs+nr-r}{nrs-ns+2s}, \frac{nrs+nr-2r}{nrs-ns+3s}, \ldots, \frac{nrs+r}{nrs}$$

included in the ratio M/Λ, I assume that the fraction having ordinal number m, namely,

$$\frac{nrs + nr - mr + r}{nrs - ns + ms},$$

is equal to $[(s + 1)/s]$. I derive therefrom

$$n = m + \frac{mr - r}{s+1} \text{ so that } nt = mt + \frac{mrt - rt}{s+1}.$$

After this, it will be shown just as before that the maximal term M of the binomial $r + s$ raised to this power will be more than $c(s - 1)$ times greater than the limit Λ, and also, consequently, that the sum of all the terms between the maximal M and the limit L will be more than c times greater than the sum of all the terms beyond this limit whose number is not more than $r - 1$ times greater [than the number of terms between M and Λ]. And so we finally conclude that, upon raising the binomial $r + s$ to the power equal to the greater of two numbers,

$$mt + \frac{mst - st}{r+1} \text{ and } mt + \frac{mrt - rt}{s+1},$$

the sum of all the terms included between the limits L and Λ exceeds more than c times the sum of all the other terms extending on either side beyond these limits. The finite power possessing the desired property is thus discovered, QED.

The Main Proposition. Now follows the proposition itself for whose sake all the previous was stated and whose demonstration ensues solely from the application of the preliminary lemmas to the present undertaking. To avoid tedious circumlocution, I name the cases in which some event can happen *fecund* or *fertile*; and *sterile* those in which the same event does not occur. In the same way, I name the experiments *fecund* or *fertile* if some fertile case appears in them and *infertile* or *sterile* when we observe something sterile.

Let the number of fertile cases be to the number of sterile cases precisely or approximately as r to s; or to the number of all the cases as r to $r + s$, or as r to t so that this ratio is contained between the limits $(r + 1)/t$ and $(r - 1)/t$. It is required to show that it is possible to take such a number of experiments that it will be in any number of times (for example, in c times) more likely that the number of fertile observations will occur between these limits rather than beyond them, that is, that the ratio of the number of fertile observations to the number of all of them will be not greater than $(r + 1)/t$ and not less than $(r - 1)/t$.

Dem[*onstration*]. Suppose that the number of the available observations is nt. It is required to determine the expectation, or probability that all of them without exception will be fecund; that all of them will be such with one, with two, 3, 4, etc being sterile. Since, according to the assumtion, there are t cases in each observation, r of them fecund and s sterile, and because separate cases of one observation can be combined with separate cases of another one, and then again combined with separate cases of the third, the 4th, etc, it is easy to see that the Rule attached to the end of the notes of Proposition 12[5.4] of Part 1 [of this book] and its second corollary containing the general formula by whose means the expectation of the lack of sterile observations, $r^{nt} : t^{nt}$, of one, two, three etc sterile observations

$$\frac{nt}{1} r^{nt-1}s: t^{nt},\ \frac{nt(nt-1)}{1\cdot 2} r^{nt-2}s^2 : t^{nt},\ \frac{nt(nt-1)(nt-2)}{1\cdot 2\cdot 3} r^{nt-3}s^3: t^{nt}, \text{ etc,}$$

are here suitable.

Therefore (after rejecting the common term t^{nt}) it becomes clear that the degrees of probability, or the number of cases in which it can happen that all the experiments are fecund, or all excepting one sterile, excepting two, 3, 4 etc sterile, are expressed, respectively, by

$$r^{nt},\ \frac{nt}{1} r^{nt-1}s,\ \frac{nt(nt-1)}{1\cdot 2} r^{nt-2}s^2,\ \frac{nt(nt-1)(nt-2)}{1\cdot 2\cdot 3} r^{nt-3}s^3, \text{ etc,}$$

that is, by the terms themselves of the binomial raised to the power of nt, which were just studied in our lemmas. All the rest is now manifest. Namely, it follows from the nature of the series that the number of cases, which add nr fecund to ns sterile observations, is indeed [corresponds to] the maximal term M since, according to Lemma 3, ns terms precede, and nr terms succeed it. In the same way, the number of cases in which there occurred either $nr + n$ or $nr - n$ fecund observations with the others being sterile, are expressed by the terms L and Λ, n terms apart on either side from the maximal term M. Consequently, the total number of cases in which there are not more than $nr + n$, and not less than $nr - n$ fecund observations, is expressed by the sum of the terms situated between the limits L and Λ. The total number of the other cases in which there occur either more or less fecund observations is expressed by the sum of the other terms beyond the limits L and [or] Λ. The power of the binomial may be taken so great that, according to Lemmas 4 and 5, the sum of the terms between the limits L and Λ inclusive is more than c times greater than the sum of all the other terms exceeding these limits. It is thus possible to take so many observations, that the number of cases in which the ratio of the number of fecund observations to the number of all of them does not exceed the limits

$$\frac{nr+n}{nt} \text{ and } \frac{nr-n}{nt} \text{ or } \frac{r+1}{t} \text{ and } \frac{r-1}{t},$$

is greater than c times the number of the other cases. That is, it will become greater than c times more probable that the ratio of the number of fecund observations to the number of all of them is contained between the limits [$(r + 1)/t$] and [$(r - 1)/t$] rather than beyond them. Quod demonstrandum erat.

When applying this to separate numerical examples, it is self-evident that the greater, in the same ratio, we assume the numbers r, s and t, the narrower can be made the boundaries $(r + 1)/t$ and $(r - 1)/t$ of the ratio r/t. Therefore, if the ratio of the number of cases r/s that should be determined by observation is, for ex[ample], one and a half, I take for r and s not 3 and 2, but 30 and 20, or 300 and 200, etc. It is sufficient to assume $r = 30$, $s = 20$ and $t = 50$ for the limits to become $(r + 1)/t = 31/50$ and $(t - 1)/t = 29/50$. Suppose in addition that $c = 1000$. Then, according to what was prescribed in the Explanatory

Comment, it will occur that, for the terms *on the left* and *on the right* respectively,

$$m > \frac{\text{Log}\,[c(s-1)]}{\text{Log}\,(r+1) - \text{Log}\,r} = \frac{42\,787\,536}{142\,405} < 301,\ nt = mt + \frac{mst - st}{r+1} < 24\,728,$$

$$m > \frac{\text{Log}\,[c(r-1)]}{\text{Log}\,(s+1) - \text{Log}\,s} = \frac{44\,623\,980}{211\,893} < 211,\ nt = mt + \frac{mrt - rt}{s+1} = 25\,550.$$

From which, as it was demonstrated there, it will follow that, having made 25 550 experiments, it will be more than a thousand times more likely that the ratio of the number of obtained fertile observations to their total number is contained within the limits 31/50 and 29/50 rather than beyond them. And in the same way, assuming c = 10 000 or c = 100 000 etc, we will find that the same is more than ten thousand times more probable if 31 258 experiments will be made; and more than a hundred thousand times if 36 966 experiments will be made; and so on until infinity, always adding 5708 other experiments to the 25 550 of them. This, finally, causes the apparently singular corollary: if observations of all events be continued for the entire infinity (with probability finally turning into complete certitude), it will be noticed that everything in the world is governed by precise ratios and a constant law of changes, so that even in things to the highest degree casual and fortuitous we would be compelled to admit as though some necessity and, I may say, fate[5.5]. I do not know whether Plato himself had this in mind in his doctrine on the restoration of all things according to which everything will revert after an innumerable number of centuries to its previous state.

Notes

2.1. It was Bortkiewicz (1917, p. x) who noticed the new word in the *Ars Conjectandi* and put it into scientific circulation, although Prevost & Lhuilier (1799, p. 3) had preceded him. The *Oxford English Dictionary* included this word, which had already appeared in ancient Greece (Hagstroem 1940), with a reference to a source published in 1662. I am not sure that the noun *stochastics* (in my translation) is generally used.

2.2. Although an astrologer, Kepler (1610, §115; p. 238 in 1941) simply refused to answer definitely the same question. Times had changed! Bernoulli resumed this discussion in his Chapter 3.

2.3. The application of stochastic reasoning to one single case conforms to modern ideas.

2.4. John Owen (1563 – 1622). Haussner (Bernoulli 1713, German transl., p. 311) saw five editions of his *Epigrams*.

3.1. Lambert indicated that the main formula above was faulty, see Haussner (Bernoulli, 1899, p. 311).

4.1. This is the very old principle of indifference. It can be perceived, for example, in the use of the arithmetic mean in astronomy since Kepler's time.

4.2. Arnauld was the main author of *L'art de penser* (Arnauld & Nicole 1662).

4.3. Bernoulli wrote *stones*; the German translation mentioned *small stones* (Steinchen).

4.4. The maximal term of the binomial $(r + s)^n$ is approximately equal to $1/\sqrt{2\pi nrs}$ and therefore decreases with an increasing n as $1/\sqrt{n}$, see e.g. Feller (1950, §3 of Chapter 6).

4.5. Adriaan Metius (1571 – 1635); Ludolph van Ceulen (1540 – 1610).

5.1. A misprint in this ratio was corrected without comment in all the translations.

5.2. Separate parts of Bernoulli's *Tractatus de Seriebus Infinitis* appeared in 1689 – 1704, and, for the first time as a single entity, in 1713 together with the *Ars Conjectandi*.

5.3. The "second" (repeated in the same sense in the Explanatory Comment below) is unusual: Bernoulli actually had in mind the term immediately neighboring M. Cf.: February is the second month of the year, not the second after January. A similar remark is of course valid with respect to the "third" and the "fourth".

5.4. Bernoulli wrongly referred to Proposition 13. Haussner (Bernoulli 1713, German transl., p. 262) corrected him without comment.

5.5. Bernoulli obviously had in mind the archaic notion of the Great Year ("innumerable number of centuries").

References

Jakob Bernoulli

(manuscript, ca. 1684 – 1690; partial publication 1975), Meditationes [Diary]. In Bernoulli (1975, pp. 21 – 90).

(1713, in Latin), *Ars conjectandi*. Reprint: J. Bernoulli (1975, pp. 107 – 259).

(1899), German translation of same by R. Haussner: *Wahrscheinlichkeitsrechnung*. Reprint: Frankfurt/Main, 1999.

(1913). Russian translation of Part 4 of same by Ya.V. Uspensky with Foreword by A.A. Markov. Reprint: Bernoulli (1986, pp. 23 – 59).

(1966), *Translations from James Bernoulli* by Bing Sung this being an English translation of Part 4 of the *Ars Conjectandi*. Dept of Statistics, Harvard Univ., Techn. Rept No. 2. The translation is very loose and therefore almost worthless.

(1975), *Werke*, Bd. 3. Basel. Ed., B.L. van der Waerden. In addition to the *Ars Conjectandi*, includes reprints of several classical contributions and commentaries.

(1986), *O Zakone Bolshikh Chisel* (On the Law of Large Numbers). Includes reprint of J. Bernoulli (1913) with comments (O.B. Sheynin, A.V. Prokhorov, N.G. Gamkrelidze), three commentaries (Sheynin, Bernoulli and the beginnings of the theory of probability; Yu.V. Prokhorov, The law of large numbers and estimation of the probabilities of large deviations; A.P. Youshkevich, Biography of Bernoulli), all this preceded by A.N. Kolmogorov's Foreword and Markov's speech of 1913 at the session of the Petersburg Academy of Sciences observing the bicentenary of the law of large numbers. Ed., Yu.V. Prokhorov. Moscow.

(1987), *Jacques Bernoulli & l'ars conjectandi*, this being a Latin-French edition of Part 4 of the *Ars Conjectandi*. Translation by B. Lalande. Ed., N. Meusnier. Paris.

Other Authors

Arnauld, A., Nicole, P. (1662), *L'art de penser.* (Appeared anonymously.) Paris, 1992.
Bernoulli, N. (1709), *De usu artis conjectandi in jure*. In J. Bernoulli (1975, pp. 284 – 326).
Bortkiewicz, L. (1917), *Die Iterationen*. Berlin.
--- (1923), Wahrscheinlichkeit und statistische Forschung nach Keynes. *Nord. Stat. Tidskr.*, t. 2, pp. 1 – 23.
David, Florence Nightingale (1962), *Games, Gods and Gambling.* London.
De Moivre, A. (1712, in Latin), De mensura sortis, or, The measurement of chance. *Intern. Stat. Rev.*, vol. 52, 1984, pp. 236 – 262. Preceded by commentary by A. Hald (pp. 229 – 236).
--- (1756), *Doctrine of Chances.* New York, 1967. Previous editions: 1718, 1738.
Descartes, R. (1644), *Les principes de la philosophie. Oeuvr.*, t. 9, pt. 2 (the whole issue). Paris, 1978, this being a reprint of the edition of 1647.
Feller, W. (1950), *Introduction to Probability Theory and Its Applications*, vol. 1. New York, 1957.
Franklin, J. (2001), *The Science of Conjecture*. Baltimore.
Garber, D., Zabell, S. (1979), On the emergence of probability. *Arch. Hist. Ex. Sci.*, vol. 21, pp. 33 – 53.
Gini, C. (1946), Gedanken zum Theorem von Bernoulli. *Schweiz. Z. Volkswirtschaft u. Statistik*, 82. Jg, pp. 401 – 413.
Graunt, J. (1662), *Natural and Political Observations Made upon the Bills of Mortality*. Baltimore, 1939. Ed., W.F. Willcox.
Hagstroem, K.-G. (1940), Stochastik, ein neues – und doch ein altes Wort. *Skand. Aktuarietidskr.*, t. 23, pp. 54 – 57.
Hald, A. (1990), *History of Probability and Statistics and Their Applications before 1750*. New York.
--- (2003), *History of the Law of Large Numbers and Consistency*. Univ. Copenhagen, Dept Applied Math. & Statistics, Preprint No. 2.
Halley, E. (1694), *An Estimate of the Degree of Mortality of Mankind.* Baltimore, 1942.
Halperin, T. (1988), The development of probability logic from Leibniz to MacColl. *Hist. and Phil. of Logic*, vol. 9, pp. 131 – 191.
Haushofer, D.M. (1872), *Lehr- und Handbuch der Statistik.* Wien.
Huygens, C. (1657), Le calcul dans les jeux de hasard. *Oeuvr. Compl.*, t. 14. La Haye, 1920, pp. 49 – 91.
Keynes, J.M. (1921), *Treatise on Probability. Coll. Writings*, vol. 8. London, 1973.
Kepler, J. (1596, 1621, in Latin), *Weltgeheimnis*. Augsburg, 1923. English transl.: New York, 1981.
--- (1610), Tertius interveniens. *Ges. Werke*, Bd. 4. München, 1941, pp. 149 – 258.
Knapp, G.F. (1872), Quetelet als Theoretiker. *Jahrb. National-Ökon. u. Statistik*, Bd. 18, pp. 89 – 124. Reprinted in author's *Einführung in einige Hauptgebiete der Nationalökonomie.* München, 1925, pp. 17 – 53.
Kohli, K. (1975), Aus dem Briefwechsel zwischen Leibniz und Jakob Bernoulli. In J. Bernoulli (1975, pp. 509 – 513).
Koopman, B.O. (1940), The bases of probability. *Bull. Amer. Math. Soc.*, vol. 46, pp. 763 – 774.

Leibniz, G.W. (Manuscript 1686), Allgemeine Untersuchungen über die Analyse der Begriffe und wahren Sätze. In author's book *Fragmente zur Logik.* Berlin, 1960, pp. 241 – 303.

Maciejewski, C. (1911), *Nouveaux fondements de la théorie de la statistique.* Paris.

Markov, A.A. (1924), *Ischislenie Veroiatnostei* (Calculus of probability). Moscow. Fourth, posthumous edition. Previous editions: 1900, 1908, 1913. German translation of the second edition: Leipzig – Berlin, 1912.

Mill, J.S. (1843), *System of Logic.* London, 1886.

Montmort, P.R. (1713), *Essay d' analyse sur les jeux de hazard*. Published anonymously. New York, 1980. First edition, 1708.

Ondar, Kh.O., Editor (1977, in Russian), *Correspondence between Markov and Chuprov*. New York, 1981.

Ore, O. (1963), *Cardano, the Gambling Scholar*. Princeton.

Oresme, N. (1966), *De proportionibus proportionum* and *Ad pauca respicientes*. Ed. E. Grant. Madison. Latin – English edition.

Pearson, K. (1925), James Bernoulli's theorem. *Biometrika*, vol. 17, pp. 201 – 210.

Prevost [P.], Lhuilier [S.A.J.] (1799), Sur l'art d'estimer la probabilité des causes par les effets. *Mém. Acad. Roy. Sci. et Belles-Lettres Berlin avec l'Histoire* 1796, pp. 3 – 24 of the second paging.

Romanovsky, V.I. (1961), *Matematicheskaia Statistika* (Mathematical Statistics), Book 1. Ed. T.A. Sarymsakov. Tashkent.

Shafer, G. (1978), Non-additive probabilities in the work of [J.] Bernoulli and Lambert. *Arch. Hist. Ex. Sci.*, vol. 19, pp. 309 – 370.

Sheynin, O. (1970), On the early history of the law of large numbers. In *Studies in the History of Statistics and Probability* [, vol. 1]. Editors, E.S. Pearson, M.G. Kendall. London, 1970, pp. 231 – 239. Reprinted from *Biometrika*, vol. 55, 1968, pp. 459 – 467.

--- (1973), Mathematical treatment of astronomical observations. *Arch. Hist. Ex. Sci.*, vol. 11, pp. 97 – 126.

--- (1977), Early history of the theory of probability. Ibidem, vol. 17, pp. 201 – 259.

--- (1982), On the history of medical statistics. Ibidem, vol. 26, pp. 241 – 286.

--- (1991), On the notion of randomness from Aristotle to Poincaré. *Math. Inform. et Sci. Humaines*, No. 114, pp. 41 – 55.

--- (2001), Social statistics and probability theory in the 19th century. *Historia Scientiarum*, vol. 11, pp. 86 – 111. Shorter version: *Jahrb. National-Ökon. u. Statistik*, Bd. 223, 2003, pp. 91 – 112.

--- (2003), On the history of the Bayes theorem. *Math. Scientist*, vol. 28, pp. 37 – 42.

Sylla, E.D. (1998), The emergence of mathematical probability from the perspective of the Leibniz – Jakob Bernoulli correspondence. *Perspectives of Sci.*, vol. 6, pp. 41 – 76.

Youshkevich, A.P. (1986, in Russian), N. Bernoulli and the publication of J. Bernoulli's Ars Conjectandi. *Theory of Probability and Its Applications*, vol. 31, 1987, pp. 286 – 303.