TRANSLATIONS FROM JAMES BERNOULLI

BING SUNG
WITH A PREFACE BY A. P. DEMPSTER
DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY

1. PREFACE

The following translations were undertaken to provide easy access to a landmark in the historical development of the ideas of probability and statistical inference. First, the fourth part of *Ars Conjectandi* by James Bernoulli (1713) is presented in its entirety, and, second, excerpts are presented from correspondence between Leibnitz and Bernoulli, as found in Leibnitz (1855). The excerpts date from 1703 and 1704, not long before Bernoulli's death on 16 August 1705. In them, Bernoulli describes the main mathematical result of the *Pars Quarta* of *Ars Conjectandi* and debates with Leibnitz its implications. The correspondence also suggests that Bernoulli was dissatisfied with his mathematical demonstration in *Pars Quarta* and indeed this was the reason for the postponement of publication until after his death.

In contemporary notation, Bernoulli shows that, if

$$P_N = \sum \binom{n}{x} p^x (1-p)^{N-x},$$

summed over

$$\left| \frac{x}{N} - [(N+1)p] \right| \leq \frac{1}{t},$$

then

$$\lim_{N \to \infty} P_N = 1$$

for fixed p and t. He also provided a lower bound for P_N sufficient to enable him to compute a value of N large enough to ensure a P_N as close to unity as desired. A mathematician of Bernoulli's stature would surely have recognized the crudeness of the lower bound for P_N, and thus it is a plausible conjecture that it was his inability to compute P_N more accurately which led Bernoulli to postpone publication. In his introduction to the posthumous *Ars Conjectandi*, James' nephew Nicholas admits his own inability to perfect the work and calls on Montmort and de Moivre to direct their energies thereto (see David, (1962), p. 134). As is well-known, de Moivre succeeded in providing the normal approximation to P_N where $p = \frac{1}{2}$ and Laplace achieved the case of general p (see Todhunter (1865)).

Considering the background and tools at his disposal, it was a remarkable achievement for Bernoulli to formulate and demonstrate a version of the law of large numbers, and then to formulate even without solution a problem whose solution led to the central limit

---

Date: February 12, 1966.

Technical Report No. 2. Contract Nonr 1866(37), NR-042-097. This research was supported by the Office of Naval Research.

theorem. It seems unfair in this light to berate Bernoulli, as did Pearson (1925), for the crudeness of his lower bound.

The early chapters of *Pars Quarta* give Bernoulli's general attitudes to what probabilities are (degrees of certainty) and how they are to be found. His exposition of this treacherous topic is vigorous, original and colorful. In the final analysis, he found himself unable in most real world situations to enumerate and weight all the proofs or cases for and against any outcome. Thus he was led to formulate the idea of repeated trials for learning *a posteriori* what is hidden *a priori*,and thence he came to his theorem.

Bernoulli scored important firsts in the history of statistical inference. He clearly separated the notion of sample value from that of population value, for he explicitly set out to infer the latter from the former. At the same time, he wrote down the sampling distribution of the sample frequency given the population frequency and clearly regarded this distribution and the theorem based on it to be fundamental for inference. Thus Bernoulli ought to be regarded as the founder of the broad school of inference, including the approaches of Fisher and of Neyman and Pearson, which attempts to make inferences on the basis of sampling distributions.

Although Bernoulli's theorem is expressed in terms of a known population frequency p, its use was clearly intended for the situation where p is unknown while a sample frequency $\hat{p}$ is known. In present day terms, the question is: does the validity of the statement $\Pr(|\hat{p}-p| < \delta) > 1-\epsilon$ before sampling with given p also carry with it the validity of the same statement after sampling with $\hat{p}$ known and p unknown? Fisher's fiducial argument answers this type of question in the affirmative while Neyman's confidence argument side-steps the question. Bernoulli did not explicitly formulate the question, and he apparently regarded the appeal of his theorem to be self-evident, much as many people now regard the appeal of the confidence argument to be self-evident.

Both Todhunter (1865) and David (1962) point out that, in the correspondence translated below, Leibnitz queried Bernoulli's inverse use of his theorem. It needs to be made clear, however, that Leibnitz was not raising the technical issue of the previous paragraph. Rather he presented a general argument against any method of trying to determine the characteristics of an infinite population from a finite sample. I find the Leibnitz argument worth pondering even today. And, of course, the issue raised implicitly by Bernoulli is very much alive and underlies the contemporary search for good methods of statistical reasoning.

2. PART FOUR OF THE ART OF CONJECTURING SHOWING THE USE AND APPLICATION OF THE PRECEDING TREATISE IN CIVIL, MORAL, AND ECONOMIC AFFAIRS

CHAPTER I

Certain preliminaries about the Certainty, Probability, Necessity, and Contingency of Events

The certainty of any event is observed either *objectively* and in itself, and it does not signify anything other than the very truth of the existence or the future existence of that event; or the certainty is observed *subjectively* and according to ourselves, and it lies in the measure of our knowledge about this truth of present or future existence.

All things which exist or are acted upon under the sun — past, present, or future things — always have the greatest certainty in themselves and objectively. This is true about present and past things: since by the very fact that they are or have been, they cannot be or have been nonexistent. Nor should it be doubted in regard to future things which, even if not by the inevitable necessity of some fate, nevertheless must happen in the future according to the plan of a sometime divine supervision and a sometime divine predetermination; for unless whatever lies in the future will occur for certain, it is not clear in what way the praise of the omniscience and omnipotence of the greatest Creator can remain undiminished. In addition, some people argue about how this certainty of future existence can agree with dependency upon or independency of secondary causes; but since this issue has nothing to do with our goal, we do not wish to touch on it.

A judgment of certainty made according to ourselves is not the same for all things but varies in many ways, sometimes more and sometimes less. Consider those things upon which it is so agreed that we can in no way doubt their existence or future existence (this may happen through revelation, understanding, perception, experience, one's own observation, or some other way): such things delight in the greatest and absolute certainty. All other things have a less perfect measure of certainty in our minds, greater or smaller according as the probabilities — which urge that the thing exists, will exist, or has existed — are greater or smaller.

For *probability* is a degree of certainty and differs from absolute certainty as a part differs from the whole. If, for example, the whole and absolute certainty — which we designate by the letter a or by the unity symbol 1 — is supposed to consist of five probabilities or parts, three of which stand for the existence or future existence of some event, the remaining two standing against its existence or future existence, this event is said to have $3/5\ a$ or $3/5$ certainty.

Therefore, a thing which has a larger part of certainty is called *more probable* than another thing; although only that thing whose probability notably exceeds half the certainty is called, out of common expression, positively *probable*. I say *notably*; for a thing whose probability is about half the certainty is called *dubious* or doubtful. And so, that which has $1/5$ certainty is more probable that that which has $1/10$ certainty, although neither is positively probable.

That which has a very little part of certainty is *possible*; that which has no or infinitely little certainty is impossible. Thus, that which has $1/20$ or $1/30$ certainty is possible.

That is *morally certain* whose probability nearly equals the whole certainty, so that a morally certain event cannot be perceived not to happen: on the other hand, that is *morally impossible* which has merely as much probability as renders the certainty of failure moral certainty. Thus, if one thing is considered morally certain which has $999/1000$ certainty, another thing will be morally impossible which has only $1/1000$ certainty.

That is *necessary* which cannot fail to exist in the present, to exist in the future, or to have existed in the past. This is brought about either by *physical necessity*: in this way it is necessary for fire to burn, for a triangle to have three angles equal to two right angles, for a full moon which occurs while the moon is at a nodal point to be subject to eclipse; or this is brought about by *hypothetical necessity*, by which every single thing — as long as it exists or has existed, or is supposed to exist or to have existed — must exist or have existed: in this sense it is necessary that Peter writes, Peter whom I know and suppose to write; or finally, this is brought about by the necessity of *agreement* or *understanding*: in this way a dice-thrower who has thrown a six with the die is said to win necessarily if it had been so previously agreed among the players that victory depends on the throw of a six.

A *contingent* thing (both an *independent* thing which relies on the judgment of a rational being and a *fortuitous* or *lucky* thing which relies on chance or fortune) is something which may not exist now, not exist in the future, or not have existed in the past; I mean things which have a remote but not imminent possibility of occurring: for contingency does not always completely exclude necessity, even with regard to secondary causes. I explain this statement in the following examples: It is most certain that when the die's position, velocity, distance from the dice-table, and momentum with which it left the hand of the thrower have been given, the die cannot fall otherwise than it actually does fall; likewise, it is most certain that when the present constituency of the air has been given, when the mass, position, movement, direction, and velocity of the winds, vapors, and clouds have been given, and when the laws of mechanics have been given — laws according to which all these things mutually affect each other in turn — the storm of tomorrow cannot occur in any other way than the way in which it will actually occur; so that these effects follow from their own closely related causes with no less necessity than the phenomena of eclipses follow from the movement of the luminary bodies: and nevertheless, experience has maintained that eclipses alone are calculated by necessities, but that the fall of the die and the future occurrence of a storm are calculated by contingencies. There is no reason for this fact other than that those data which are supposed to determine later events (and especially such data which are in nature) have nevertheless not been learned well enough by us. If they have been learned well enough by us, the studies of geometry and physics have not been well enough refined so that these effects can be calculated from data in the same way that eclipses can be computed and predicted from the observed principles of astronomy. Therefore, just as much as the fall of a die and the occurrence of a storm, these very eclipses — before astronomy had been advanced to such a degree of perfection — needed to be referred among future contingencies. Hence it follows that what can seem to be to one person at one time a contingent event may be at another time to another person (indeed, to the very same person) a necessary event after its causes have been learned; so that contingency especially may also depend upon our knowledge, in so far as we see no objection for a thing's non-existence or non-future-existence although here and now, by dint of a cause close but unknown to us, it exists or may happen out of stark necessity.

Good fortune (*un Bonheur*, *ein Glück*) and adverse fortune (*un Malheur*, *ein Unglück*) are said to occur when good or bad happens to us: not any at all good or bad, but a good or bad which was *more* likely — or at least just as likely — *not* to have happened to us. And so, fortune is the better or the worse the *less* likely it was that the event would turn out to be good or bad.

Thus, he is remarkably fortunate who has found a treasure by digging up the earth, because this does not happen once in digging a thousand times. If twenty deserters —

one of whom must be killed by hanging as an example for the others — vie for life in a game of chance, those nineteen are not properly called fortunate who have obtained a kinder fate; but rather the twentieth is called most unfortunate to whom the black fate has fallen. And so, your friend who has escaped unharmed from a battle in which a very small part of the battlers dies should not be declared fortunate, unless, perhaps, you think that he should be called fortunate because he was singled out by the good luck which is basic to the preservation of life.

CHAPTER II

About Science & Conjecture, About the Art of Conjecturing. About Proofs of Conjectures. Certain General Axioms Pertaining Hereto.

We are said to *know* or to *understand* those things which are certain and beyond doubt; all other things we are said merely to *conjecture* or *guess about*.

To *conjecture about* something is to measure its probability; and therefore, the *art of conjecturing* or the *stochastic art* is defined by us as the art of measuring as exactly as possible the probabilities of things with this end in mind: that in our decisions or actions we may be able always to choose or to follow what has been perceived as being superior, more advantageous, safer, or better considered; in this alone lies all the wisdom of the philosopher and all the discretion of the statesman.

Probabilities are estimated by the *number* and the *weight of proofs*, proofs which in any way ascertain or show that some thing exists, will exist, or has existed. Moreover, I judge the power of a proof by its *weight*.

Proofs themselves are either *intrinsic*, usually artificial, and taken from places relating to the thing's cause, effect, subject, accessory, or sign, or relating to any other circumstance which seems to have any connection whatsoever with proving the thing; or proofs are *extrinsic*, not artificial, and are derived from the authority and testimony of men. Take this example: Titius is found dead on the road and Maevius is accused of committing the murder; the proofs for the accusation are these:

(1) That it is well-known that Maevius regarded Titius with hatred (a proof for *cause*, for this hatred itself could have driven Maevius to kill.)
(2) That upon being interrogated, Maevius turned pale and answered apprehensively (a proof for *effect*; for paleness and fear itself could have proceeded from his own cognizance of having committed a crime.)
(3) That a bloodstained sword was found in Maevius' house (sign).
(4) That on the day that Titius was slain on the road, Maevius traveled over that same road (*circumstance* of place and time.)
(5) And finally, that Gaius alleges that on the day before the murder he had interceded in a dispute between Titius and Maevius (*testimony*.)

Moreover, before we draw nearer to our purpose — which is to show how one should use these proofs for conjectures in measuring probabilities — it is desirable to set forth some general rules or axioms which simple thought is wont to dictate to any man of sound mind, and which are also continually observed by more judicious men in the experience of civil life.

1. *There must not be a place for conjectures in things about which one may attain complete certainty*. Therefore, it would be useless if an astronomer, from the fact that he perceived eclipses occurring two or three times a year, chose to prophesy whether or not some full moon is subject to eclipse, for one can obtain the truth of the matter by means of definite calculation. And so, if a thief, upon being interrogated, answered that he sold the

stolen object to Sempronius, the case would be handled foolishly by a judge who—from the expression and tone of the man as he talked, or from the quality of the object stolen in the theft, or from some other circumstances of the theft—chooses to conjecture about the probability of the statement even when Sempronius is at hand; for the judge will be able to learn from Sempronius about the whole matter with certainty and with ease.

2. *It is not enough to weigh one or another proof, but everything must be sought out which can come within our realm of knowledge and which appears to have any connection at all with proving the thing*. For example: three ships set sail from port; after some time it is announced that one of them suffered shipwreck; which one is guessed to be the one that was destroyed? If I considered merely the number of ships, I would conclude that the misfortune could have happened to each of them with equal chance; but because I remember that one of them had been eaten away by rot and old age more than the others, had been badly equipped with masts and sails, and had been commanded by a new and inexperienced captain, I consider that this ship, more probably than the others, was the one to perish.

3. *One must attend not only to those things which serve to prove the thing, but also to all those things which can be adduced to prove the opposite of the thing, so that after both sides are weighed it will be clear which of them has more weight*. It is asked about a friend who has been away from his homeland for a very long time whether he can be declared dead. These proofs serve for the affirmative: that although every effort has been made, nothing has become known of him for all of twenty years; that men who wander about are exposed to a great many perils of life, perils from which men who remain at home are exempt; that therefore he may have ended his life in the sea, he may have died on the road, he may have been killed in battle, he may have perished from disease or misfortune in some place where he was known by no one; that if he were among the living, he must now be of that age which few men reach even if they have lived at home; that he would have written even if he dwelled along the outermost regions of India because he knew that an inheritance awaited him at home; and there are other proofs. However, one must not acquiesce to these proofs, but rather the following proofs, which support the negative, must be set forth in opposition: It is well-known that the man was lazy, took hold of the pen painfully, and held his friends in contempt; perhaps he was led away as a captive of the Barbarians so that he was not able to write; perhaps also he wrote several times from India, but the letters were lost due either to the carelessness of the letter-carriers or to the shipwreck of the mailboat; finally, it is well-known that many people have remained away for a longer time and nevertheless have finally returned safe and sound.

4. *Remote and general proofs are sufficient for judging about general events; but for forming conjectures about specific events, more closely related and special proofs must be added, if only they can be obtained*. And so, when it is asked abstractly how much *more* probable it is that a young man of twenty years will outlive an old man of sixty years *than* that the latter will outlive the former, there is nothing besides the difference of age and years which you can consider; but when the conversation is particularly about the specific young man Peter and the specific old man Paul, one must attend moreover to their individual complexions and the concern with which each looks after his own health. For if Peter is sickly, if he indulges in his fancies, if he lives intemperately, it can happen that Paul, although more advanced in age, may nevertheless most reasonably be able to foster hope for a longer life.

5. *In uncertain and dubious matters our actions should be suspended until more light has come to bear upon the matter; but if the occasion allows no delay of action, the choice*

that must be made from two alternatives is that which seems more suitable, safer, wiser, or more probable, although neither choice is positively suitable, safe, wise or probable. Thus, in a fire which has arisen and from which you cannot otherwise escape except by leaping headlong either from the top of the roof or from a lower story of the building, it will prove better to choose the latter because it is safer, although leaping from either place is neither purely safe nor effected without the danger of injury.

6. *That which can be useful on some occasion and harmful on no occasion is to be preferred to that which is useful on no occasion and harmful on no occasion*: therefore, it aims at this statement which I quote in our vernacular: *hilft es nicht/ so schadet es nicht*.[1] This sentence follows immediately from the preceding one; for what can be useful is better, safer, and more desirable — with other things being equal — than that which cannot be useful.

7. *One must not decide about the value of human actions from their outcomes*; since the most stupid actions sometimes enjoy the best success, and, on the other hand, the most prudent actions occasionally enjoy the worst success. Hence, the poet writes: *I hope that he is without success who thinks that deeds must be judged by their outcomes*. And so, if someone undertakes to throw three sixes with three dice on his first turn, even if he perhaps wins, he is still considered to have acted stupidly. One must censure the perverse judgments of the common rabble by whom a man is considered more excellent the more fortunate he is; nay, as far as the common rabble are concerned, a crime which brings prosperity and happiness is usually called a virtue. Ovennus wrote elegantly about this:

For that which turned out happily, though it was ill-advised,
 Ancus is considered wise, Ancus who was merely a fool;
For that which had been wisely foreseen, if it turns out badly,
 Cato himself will be a fool in the judgment of the people.

The Book of Epigrams, 11. 216 ff.

8. *In our judgments we must beware lest we attribute to things more than is fitting to attribute, lest we consider something which is more probable than other things to be absolutely certain, and lest we foist this more probable thing upon other people as something absolutely certain*. For one must see to it that the credibility which we attribute to things is proportional to the degree of certainty which every single thing possesses and which is less than absolute certainty to the same degree that the probability of the thing is less than 1; as we put it in our vernacular: *Man muß ein jedes in seinem Wert und Unwert beruhen lassen*.[2]

9. *Because it is still rarely possible to obtain total certainty, necessity and use desire that what is merely morally certain be regarded as absolutely certain*. Hence, it would be useful if, by the authority of the magistracy, limits were set up and fixed concerning moral certainty. I mean, if it were fixed whether $99/100$ certainty would suffice for producing moral certainty, or whether $999/1000$ certainty would be required. Note that then a judge could not be biased, but he would have a guideline which he would continually observe in passing judgment.

Everyone educated by the daily experience of things can, by his own exertion, pound out still more axioms of this kind; we can hardly remember all of them, however, outside of a given situation.

[1] If it doesn't help, it doesn't harm.

[2] One must let each thing lie in its worth and worthlessness.

CHAPTER III
About Various Kinds of Proofs and How Their Weights Should be Valued For Reckoning The Probabilities of Things

He who examines various proofs upon which an opinion or conjecture is based must notice a threefold difference among them: for some of them *necessarily exist and indicate the thing contingently*; others *exist contingently and indicate the thing necessarily*; and still others *both exist and indicate the thing contingently*. I clarify this difference in examples: My brother has not sent a letter to me for a long time; I wonder whether sloth or business are to blame; I also am afraid that he may perhaps be dead. Here there are three proofs for the interrupted writing: *Sloth*, *Death*, and *Business*. The first of these exists necessarily—by hypothetical necessity, since I know and assume my brother to be lazy—but indicates the interrupted writing contingently; for this sloth did not necessarily prevent him from writing. The second proof exists contingently—for my brother may still be alive—but indicates the interrupted writing necessarily, since a dead man cannot write. The third proof exists contingently, for he may or may not have business, and if he has any business, it is not necessarily such that it prevents him from writing.

Another example: I take a dice-player to whom the prize must be given, according to the rule of the game, if he throws seven points with two dice, and I wish to guess how much hope he has of winning. Here the proof for victory is the throw of seven, a throw which indicates the victory necessarily (by the necessity, of course, of the agreement entered upon by the players), but exists only contingently; for other numbers of points beside seven are able to fall.

Beyond this distinction among proofs there is also another difference one should observe: some proofs are *pure* and others are *mixed*. I call those *pure* which in certain cases prove the thing in such a way that in other cases they prove nothing positively; I call those *mixed* which in some cases prove the thing in such a way that in other cases they prove the opposite of the thing. Take this example: In a mob of rowdy men a man is stabbed with a sword, and, according to the trusted testimony of worthy men who are observing the mob from a distance, it is agreed that the person who committed the crime was wearing a black cloak. If among the rowdy men Gracchus and three others are found clad in tunics of that color, this tunic is a proof that the murder was committed by Gracchus, but it is mixed: since in one case it proves his guilt and in three cases it proves his innocence, in that, obviously, the murder was perpetrated either by him or by one of the other three men, and the murder could not have been perpetrated by one of the three men unless Gracchus is supposed innocent of the crime. But if indeed Gracchus grows pale in the subsequent hearing, this pallor of his face is a pure proof, for it proves Gracchus's guilt if it arises from a guilty conscience, but it does not in turn prove his innocence if it springs from another source: for it may happen that Gracchus grows pale for another reason but nevertheless is still the murderer.

From what has been said previously, it is clear that the power which any proof has depends upon a multitude of cases in which it can exist or not exist, in which it can indicate or not indicate the thing, or even in which it can indicate the opposite of the thing. And so, the degree of certainty or the probability which this proof generates can be computed from these cases by the method discussed in the first part just as the fates of gamblers in games of chance are accustomed to be investigated.

In order to show this, we assume that b is the number of cases for which it can happen that some proof exists; that c is the number of cases for which it can happen that this proof does not exist; and that $a = b + c$. Likewise, we assume that β is the number of cases for which it can happen that it indicates the thing; that γ is the number of cases for which it can happen that it does not indicate the thing or that it indicates its opposite; and that $\alpha = \beta + \gamma$. Moreover, I assume that all cases are equally possible, or that they all can happen with equal ease; for in other cases discretion must be applied, and any case which occurs rather readily must be counted as many times as it occurs more readily than others. For example, a case which occurs three times more readily than the other cases must be counted as three cases, each of which can occur with ease equal to that of any of the other cases.

1. And so, if a proof *exists contingently* and *indicates the thing necessarily*, there will be, from the above assertions, b cases for which it can exist and therefore indicate the thing (or 1), and c cases for which it cannot exist and so does not indicate anything. This yields (by Corollary I, Prop. III of the first part) $\frac{b\cdot 1+c\cdot 0}{a} = \frac{b}{a}$, so that such a proof proves $\frac{b}{a}$ of the thing or of the certainty of the thing.

2. Next, let the proof *exist necessarily* and *indicate the thing contingently*: there will be, by hypothesis, β cases for which it can happen that it indicates the thing, and γ cases for which it can happen that it does not indicate the thing or indicates the opposite of the thing; this now makes the force of the proof for proving the thing $\frac{\beta\cdot a+\gamma\cdot 0}{\alpha} = \frac{\beta}{\alpha}$. Therefore, a proof of this kind proves $\frac{\beta}{\alpha}$ of the certainty of the thing, and moreover if it is mixed, it proves in the same way $\frac{\gamma\cdot 1+\beta\cdot 0}{\alpha} = \frac{\gamma}{\alpha}$ of the certainty of the opposite of the thing.

3. Suppose that some proof *exists contingently and indicates the thing contingently*, and suppose I immediately assume that it exists (in which case, by what has been shown above, it is known to prove $\frac{\beta}{\alpha}$ of the thing and, if it is mixed, $\frac{\gamma}{\alpha}$ of the opposite): since there are b cases for which it can exist and c cases for which it cannot exist and hence can prove nothing, this proof will have for proving the thing the value $\frac{b\left(\frac{\beta}{\alpha}\right)+c(0)}{a} = \frac{b\beta}{a\alpha}$; and if it is mixed, it will have for proving the opposite value $\frac{b\left(\frac{\gamma}{\alpha}\right)+c(0)}{a} = \frac{b\gamma}{a\alpha}$.

4. Furthermore, if several proofs are at hand for proving the same thing and if they are as shown below:

Proof:	1st.	2nd.	3rd.	4th	5th	$\cdots$
The Total Number of Cases:	a	d	g	p	s	
Proving the Thing:	b	e	h	q	t	
Not Proving, Or Proving the Opposite:	c	f	i	r	u	

the power of the proof resulting from the union of all the proofs is estimated in the following way: First, let all the proofs be *pure*; the weight of the first proof will be separately equal to $\frac{b}{a} = \frac{a-c}{a}$ (or $\frac{\beta}{\alpha}$ if it indicates the thing contingently, or $\frac{b\beta}{a\alpha}$ if it also exists contingently.) Then, let another proof be added which proves the thing (or 1) in $e = d - f$ cases and proves nothing in f cases; in these last f cases, only the weight of the first proof, which has been shown to be $\frac{a-c}{a}$, is effective. The weight produced from both arguments will be $\frac{(d-f)\cdot 1+f\left(\frac{a-c}{a}\right)}{d} = \frac{ad-cf}{ad} = 1 - \frac{cf}{ad}$. Now, let a third proof be joined; there will be h or $g - i$ cases which prove the thing, and i cases for which there is no proof and for which only the two earlier proofs retain their power of proof $\frac{ad-cf}{ad}$. Whence the force of all three is

$$\frac{(g-i)\cdot 1 + i\left(\frac{ad-cf}{ad}\right)}{g} = \frac{adg-cfi}{adg} = 1 - \frac{cfi}{adg}.$$

And so on successively if more proofs are at hand. From this it is clear that all these proofs joined together induce a probability which is less than the absolute certainty of the thing (or unity) by that part of unity arising from the division of the product of all the non-proving cases by the product of all the cases in all the proofs.

5. Now let all the proofs be *mixed*: Since the number of proving cases of the first proof is b, of the second e, of the third h, etc., and since the number of proving-the-contrary cases is $c,\ f,\ i$, etc., the probability of the thing to the probability of the opposite of the thing—by force of the first proof alone—is as b is to c; and by force of the second proof alone is as e is to f; and by force of the third proof alone is as h is to i; and so on. Whence it is clear enough that the total force of proving the thing resulting from the union of all the proofs is made up of the forces of all the individual proofs; i.e., it is clear that the probability of the thing to the probability of the opposite of the thing is in the ratio beh etc. to cfi etc., so that the absolute probability of the thing is $\frac{beh}{beh+cfi}$, and the absolute probability of the opposite of the thing is $\frac{cfi}{beh+cfi}$.

6. Moreover, let some proofs be *pure* (e.g., the first three) and some *mixed* (e.g., two others.) First I consider only the pure ones which, by paragraph 4 above, prove $\frac{adg-cfi}{adg}$ of the certainty of the thing, lacking unity by $\frac{cfi}{adg}$. Therefore, there are, as it were, $adg-cfi$ cases in which these three proofs taken jointly prove the thing (or 1), and cfi cases in which they prove nothing and in which they therefore yield a place for proving the thing to the mixed proofs alone. Furthermore, these two mixed proofs—by the preceding paragraph 5—prove $\frac{qt}{qt+ru}$ of the thing and $\frac{ru}{qt+ru}$ of the opposite of the thing. And so the probability of the thing resulting from all the proofs is[3]

$$\frac{(adg-cfi)\cdot 1+(cfi)\left(\frac{qt}{qt+ru}\right)}{adg}=\frac{adgqt+adgru-cfiru}{adgqt+adqru}=1-\frac{cifru}{adg(qt+ru)},$$

which is lacking from complete certainty or unity by the product of $\frac{cfi}{adg}$ (the number by which the probability of the thing resulting from the pure proofs lacks unity, according to paragraph 4) and $\frac{ru}{qt+ru}$, the absolute probability of the opposite of the thing drawn from the mixed proofs, according to the preceding paragraph 5.

7. But if besides the proofs which serve to prove the thing other pure proofs offer themselves, proofs by which the opposite of the thing is advised, the proofs of both kinds must be weighed separately according to the preceding rules in order that there then may exist a ratio between the probability of the thing and the probability of the opposite of the thing. Whence it must be noted that if the proofs adduced for each side are strong enough, it can happen that the absolute probability of each side notably exceeds half the certainty; i.e., that each of the alternatives is rendered probable, although relatively speaking one is less probable than the other. And so it can happen that a thing has $2/3$ certainty and its opposite possesses $3/4$ certainty; in this way each of the alternatives will be probable, but nevertheless the thing is less probable than its opposite, and is the ratio $2/3$ to $3/4$, or 8 to 9.

I cannot conceal the fact here that in the specific application of these many rules, I foresee many things happening which can often cause one to be badly mistaken if he does not proceed cautiously in discerning proofs. For sometimes proofs can appear to be distinct when they constitute, in fact, one and the same proof; or, in turn, they seem to be one proof when they are, in fact, distinct proofs; also, in a proof such things are sometimes assumed which make the proof of the contrary impossible; and so on. As an illustration of this fact I

[3]"... This formula is inaccurate. For the supposition $q=0$ amounts to having one argument *absolutely decisive against* the conclusion, while yet the formula leaves still a certain probability *for* the conclusion." Todhunter (1865).

adduce only one or two examples: I assume in the example of Gracchus mentioned above that the men who saw the brawl are trustworthy; that these men saw the murderer had red hair; and that Gracchus was noted, along with two others, to have red hair, but that neither of the other two was garbed in a black tunic. If now, from the evidence that three others besides Gracchus were garbed in black and that two others besides Gracchus were noted with red hair, someone wished to infer by paragraph 5 above that the probability of guilt to the probability of innocence in Gracchus's case is composed of the product $1/3$ and $1/2$—i.e., is $1/6$—and hence that Gracchus is much more probably innocent than guilty, he would be inferring very wrongly. For here there are not really two proofs but only one and the same proof supported by the two simultaneous circumstances of the color of the garments and the color of the hair. These two circumstances, when joined together in the person of Gracchus alone, show for sure that no one else could have been the murderer.

Take another example: There is doubt about whether the date fixed on a certain contract was fraudulently written in advance. This can be a proof for the negative: that the document was signed by the hand of a Notary—i.e., a public and sworn person—for whom the committing of this crime is very unlikely since he could not have done it without the greatest danger to his honor and fortune; and on this account also, scarcely one out of fifty may be found who would dare to perpetrate this kind of fraud. But proofs for the affirmative can be the following: that the reputation of this Notary is spoken of very poorly; that he could expect a lot of profit from this fraud; and finally, that he swears to such things which have no probability—e.g., if he had written that someone lent another person 10,000 gold pieces at the time when this someone, in the estimation of everyone, could scarcely have had 100 gold pieces in all his worldly goods. Here, if you trust only in the proof supported by the nature of the position of the undersignee, you could think that the probability of the authenticity of the document is about $49/50$. But if you weigh the contrary proofs, you are bound to acknowledge that it could hardly happen that there is a mistake in the document; and so, it is morally certain that fraud took place in this affair; that is, it is about $999/1000$ probable. But now it must not be concluded that the probability of authenticity to the probability of fraud, by paragraph 7, is in the ratio of $49/50$ to $999/1000$, i.e., in a ratio of near equality: for by the very assumption that this Notary is of befouled reputation, I am assuming also that he cannot be considered in the number of those 49 honest Notaries who detest fraud, but rather that he is the fiftieth who does not have any scruples and who serves faithlessly in his office. In short, this assumption razes and destroys all the force of that proof which at one time could have shown the authenticity of the document.

CHAPTER IV

About the two Ways of investigating the Numbers of Cases.
How one must view the Way which is based on Trials.
The Main Problem for the Way based on Trials.
And other Matters pertaining hereto.

It has been shown in the preceding chapter — from the numbers of cases in which proofs of any given thing can exist or not exist, can indicate or not indicate or even indicate the opposite — how the strengths of these proofs and the probabilities proportional to them can be calculated and estimated. And there we concluded that for correctly forming conjectures about anything at all, nothing is required other than that the number of these cases be accurately determined and that it be found out how much more easily some cases can happen than others. But here, finally, we seem to have met our problem, since this may be done only in a very few cases and almost nowhere other than in games of chance the inventors of which, in order to provide equal chances for the players, took pains to set up so that the numbers of cases would be known and fixed for which gain or loss must follow, and so that all these cases could happen with equal ease. For in several other occurrences which depend upon either the work of nature or the judgment of men, this by no means is the situation. And so, for example, the numbers of cases in dice are known: for in each die there are clearly as many cases as there are sides, and they are all equally likely. Because of the similarity of the sides and the balanced weight of the die, there is no reason why one of the sides should be more prone to fall than another, as there would be reason if the sides were of different shapes or if the die was made of a material heavier on one side than on another. And so likewise, the numbers of chances for drawing forth a white or black pebble from an urn are known, and it is known that all chances are equally likely: for the numbers of pebbles of each kind are known and determinate, and there is no reason why this pebble or that pebble should come forth rather than any other one.

But what mortal will ever determine, for example, the number of diseases — i.e., the number of cases — which are able to seize upon the uncountable parts of the human body at any age and which can inflict death upon us? And what mortal will ever determine how much more likely this disease than that disease, pestilence than dropsy, dropsy than fever will destroy a man so that then a conjecture can be formed about the relationship between life and death in future generations? Who likewise will reckon up the innumerable cases of mutations to which the air is daily exposed, so that he can then guess after any given month, not to mention after any given year, what the constitution of the air will be? Again, who has well enough examined the nature of the human mind or the amazing structure of our body so that in games which depend wholly or in part on the acumen of the former or the agility of the latter, he could dare to determine the cases in which this player or that can win or lose? For these and other such things depend upon causes completely hidden from us, and since moreover these things will forever deceive our effort because of the innumerable variety of their combinations, it would clearly be unwise to wish to learn anything in this way.

But indeed, another way is open to us here by which we may obtain what is sought; and what you cannot deduce *a priori*, you can at least deduce *a posteriori* — i.e., you will be able to make a deduction from many observed outcomes of similar events. For it must be presumed that every single thing is able to happen and not to happen in as many cases as it was previously observed to have happened and not to have happened in like circumstances. For if, for example, an experiment was once conducted on 300 men of the age and constitution of which Titius is now, and you observed that 200 of them had died before passing the next ten years and that the others had further prolonged their lives, you could safely enough conclude that the number of cases in which Titius must pay his debt to nature within the next ten years is twice the number of cases in which he can pay his debt after ten years. And so, if anyone has observed the weather for the past several years and has noted how many times it was calm or rainy; or if anyone has judiciously watched two players and has seen how many times this one or that one has emerged victorious: in this way he has detected what the ratio probably is between the number of cases in which the same events, with similar circumstances prevailing, are able to happen and not to happen later on.

And this empirical way of determining the numbers of cases by trials is neither new nor unusual, for the celebrated author of the *Ars Cogitandi*,[4] a man of great insight and intelligence, prescribes a similar method in Chapter 12 & ff. of the last part of this work;

[4]The author of this work which first appeared in 1664 in Paris is anonymous. [This is not true. The author is Arnauld.–RJP]

and everyone constantly observes the same thing in daily life. Further, it cannot escape anyone that for judging in this way about any event at all, it is not enough to use one or two trials, but rather a great number of trials is required. And sometimes the stupidest man — by some instinct of nature *per se* and by no previous instruction (this is truly amazing) — knows for sure that the more observations of this sort that are taken, the less danger will be of straying from the mark. Moreover, although this fact is known naturally to all, a proof in which this is evinced from first principles is by no means obvious, and therefore it is incumbent upon me to give such a proof here, since I think I would be doing too poor a job if I stood still in proving this one thing of which no one is aware.

Something further must be contemplated here which perhaps no one has thought about till now. It certainly remains to be inquired whether after the number of observations has been increased, the probability is increased of attaining the true ratio between the numbers of cases in which some event can happen and in which it cannot happen, so that this probability finally exceeds any given degree of certainty; or whether the problem has, so to speak, its own asymptote — that is, whether some degree of certainty is given which one can never exceed, so that however many observations are made, we can never be more than $1/2$ or $2/3$ or $3/4$ certain that we have detected the true ratio of the cases. So that what I mean may be seen in an example, I assume that in a certain urn there are hidden — without your knowledge — three thousand white pebbles and two thousand black ones, and that you, in order to investigate their numbers by experiments, draw one pebble after another (replacing each time that pebble which you drew out before you draw the next one, so that the total number of pebbles in the urn is not diminished), and that you observe how often a white pebble comes out and how often a black one comes out. It is asked whether you can do this so many times that it will be ten times, a hundred times, a thousand times, etc., *more* probable that the numbers of times in which you choose a white pebble and in which you choose a black pebble will have the same $3 : 2$ ratio which the numbers of white and black pebbles actually have, *than* that these numbers will have any other ratio different from $3 : 2$. For if this is not the case, I admit that we will be through with our attempt at exploring the numbers of cases by trials. But if this is the case and if, finally, moral certainty is reached in this way (I will show in the following chapter how this is actually done), we will have investigated the numbers of cases *a posteriori* almost as accurately as if they had been known to us *a priori*.

And according to Axiom 9 Chapter II, this is amply sufficient in civil life, where what is morally certain is considered as absolutely certain in order to form our conjectures in any situation that may arise no less scientifically than in games of chance. And in fact, if, for example, we replace the urn by the atmosphere or the human body (both of which contain fuel for various mutations and diseases as the urn contains pebbles), we will in the same way be able to determine by observations how much more easily this or that event can take place in the regions of the atmosphere or the human body.

However, lest these remarks be misunderstood, it must be noted that I do not wish for this ratio, which we undertake to determine by trials, to be accepted as precise and accurate (for then the contrary would result from this, and it would be the more *un*likely that the true ratio had been found the *greater* the number of observations made); but rather, the ratio is taken in some approximation: i.e., it is bounded by two limits. Moreover, these limits can be set up as close together as one wishes. Indeed, if in the example of the urn mentioned above we take two ratios — e.g., $301/200$ & $299/200$, or $3001/2000$ & $2999/2000$, etc. — one of which is a little larger, the other a little smaller than $1\frac{1}{2}$, it will be shown that for any given probability, it is more likely that the ratio obtained by frequently repeated trials will fall within these limits of $1\frac{1}{2}$ than outside these limits.

Therefore, this is the problem which I now set forth and make known after I have already pondered over it for twenty years. Both its novelty and its very great usefulness, coupled with its just as great difficulty, can exceed in weight and value all the remaining chapters of this thesis. However, before I go into its solution, let me briefly do away with objections which some learned men have placed in opposition to these ideas.

1. First, they say in objection that the ratio of pebbles is of one kind and that the ratio of diseases or mutations of the air is of another kind; that the number of pebbles is determinate but the number of diseases or mutations is indeterminate and uncertain. To this I answer that with respect to our knowledge both numbers are taken to be equally uncertain and indeterminate. But anything that is inherently uncertain and indeterminate cannot be understood by us any more than it can be understood that something has been and has not been created by God: for whatever God has made, by this very fact of creation He has also made it determinate.

2. Secondly, they say in objection that the number of pebbles is finite but the number of diseases, etc., is infinite. To this I answer that the latter number is vastly huge rather than infinite. But let us grant, for the sake of argument, that it is infinite; it is known that there also can exist between two infinite numbers a determinate ratio which is expressible by finite numbers either accurately or, at the very least, as closely as one wishes. Of such a kind, certainly, is the determinate ratio of the circumference of a circle to its diameter, a ratio which cannot be expressed accurately except by the infinitely continued decimal places of Ludolphus , but which, nevertheless, is bounded by Archimedes, Metius, and Ludolph himself[5] within limits which are very sufficient for practical application. Whence, a ratio between two infinite which is expressible as accurately as one wants by finite numbers may also be determined by finite trials.

3. Thirdly, they say that the number of diseases does not remain constantly the same, but new diseases sprout up every day. To this I answer that we cannot deny that the diseases may be multiplied by the passage of time, and it is certain that he who would wish to conclude about antediluvian times on the basis of today's observations would grossly error from the truth. But then nothing follows other than that sometimes new observations must be made, just as new observations would have to be made in the case of the pebbles if their number is assumed to be changed in the urn.

CHAPTER V

The Solution to the Preceding Problem

In order to explain this long proof with as much brevity and clarity as possible, I will try to reduce everything to abstract mathematics from which I draw forth the lemmas that follow. After these lemmas have been shown, their stark application will yield the other things.

Lemma 1: Take any series of natural numbers $0,\ 1,\ 2,\ 3,\ 4,$ etc., the last and greatest member of which is $r + s$ and an intermediate term of which is r. r is bounded on either side by $r - 1$ and $r + 1$. If this series is continued until the last term becomes a multiple of the number $r + s$ (I mean until it becomes $nr + ns$, n a natural number), and if the intermediate terms r, $r + 1$, and $r - 1$ are increased in the same way so that they become

[5]Archimedes (c. 287 – 212 BC); Adriaen Anthonisz — whose pseudonym was Metius — (1527–1607 AD); Ludolph van Ceulen (1540–1610 AD): these are the people Bernoulli is referring to.

nr, $nr+n$, and $nr-n$ respectively, the original series initially taken:

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

is changed into this series:

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

Indeed, with an increasing value of n, *the number* of terms greater than nr up to and including $nr+n$ ($=$ *the number* of terms less than nr down to and including $nr-n$) will increase, as will the number of terms greater than $nr+n$ up to an including $nr+ns$ and the number of terms less than $nr-n$ down to and including 0. However, no matter how large an n is taken, the number of terms greater than $nr+n$ up to and including $nr+ns$ will never be more than $(s-1)$ times the number of terms greater than nr up to and including $nr+n$; and also, the number of terms between $nr-n$ down to and including 0 will never be more than $(r-1)$ times the number of terms less than nr down to and including $nr-n$. For after performing a subtraction, it is clear that the number of terms greater than $nr+n$ up to and including $nr+ns$ is $ns-n$; and the number of terms less than $nr-n$ down to and including 0 is $nr-n$. And it is always true that $ns-n:n$ as $s-1:1$, and that $nr-n:n$ as $r-1:1$.[6]

Lemma 2: The number of terms in the expansion of the binomial $(r+s)$ raised to any positive integral power is one more than the value of that power. For there are three terms in the expansion of $(r+s)^2$, four terms in the expansion of $(r+s)^3$, five terms in $(r+s)^4$, etc., as is known.

Lemma 3: In any expansion of the binomial $(r+s)$ raised to a power which is a multiple of the binomial — that is, raised to a power t where $t=nr+rs$, n a natural number — if some terms precede and other terms follow a certain term M such that the number of all those preceding is in the ratio to the number of all those following as s is to r, (or if, restating this, the exponents of the r and of the s terms in M are in the ratio $r:s$), then this term M will be the largest one in the expansion, and the closer a term is to M, the larger is its value. Moreover, the ratio between this term M and another term V a given distance from M is *smaller* than the ratio between V and another term W the same distance away from V.[7]

Proof 1: It is known by mathematicians that the binomial $(r+s)$ raised to the power nt — i.e., $(r+s)^{nt}$ — is expressed by the following sum:

$$x^{nt}+\frac{nt}{1}\cdot r^{nt-1}\cdot s^1+\frac{nt(nt-1)}{1\cdot 2}r^{nt-2}s^2 \\ +\frac{nt(nt-1)(nt-2)}{1\cdot 2\cdot 3}r^{nt-3}s^3+\cdots+\frac{nt}{1}\cdot r\cdot s^{nt-1}+s^{nt};$$

in this expansion, the exponent of one part of the binomial — r — is gradually diminished, the exponent of the other part — s — is gradually increased, the coefficient of the second and the next to last term is $\frac{nt}{1}$, of the third and $(nt-1)^{\text{st}}$ term is $\frac{nt(nt-1)}{1\cdot 2}$, of the fourth and $(nt-2)^{\text{nd}}$ term is $\frac{nt(nt-1)(nt-2)}{1\cdot 2\cdot 3}$, and so on. And since, by Lemma 2 above, the number of all the terms not including M is nt $(=nr+ns)$, and since by hypothesis the ratio of the number of terms preceding M to the number of terms following M is the ratio s to r,

[6]Bernoulli is assuming the reader is aware that the number of terms greater than nr up to and including $nr+n=$ the number of terms less than nr down to and including $nr-n=n$.

[7]i.e., $(r+s)^{nr+ns} = \mathrm{r}^{nr+ns}+\binom{nr+ns}{1}r^{(nr+ns-1)}s^1+\cdots+M+\cdots+V+\cdots+W+\cdots+s^{nr+ns}\Rightarrow\frac{M}{V}<\frac{V}{W}$.

the number of terms which precede M is ns and the number of terms which follow M is nr. Whence, according to the law of binomial expansion, the term M will be:

$$\frac{nt(nt-1)(nt-2)\cdot\ldots\cdot(nr+1)r^{nr}s^{ns}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot ns},$$

or, equivalently

$$\frac{nt(nt-1)(nt-2)\cdot\ldots\cdot(ns+1)r^{nr}s^{ns}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot nr}.$$

Similarly, the terms nearest to M are:

on the left:

$$L_1=\frac{nt(nt-1)\cdot\ldots\cdot(nr+2)r^{nr+1}s^{ns-1}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot(ns-1)}$$

$$L_2=\frac{nt(nt-1)\cdot\ldots\cdot(nr+3)r^{nr+2}s^{ns-2}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot(ns-2)}$$

on the right:

$$R_1=\frac{nt(nt-1)\cdot\ldots\cdot(ns+2)r^{nr-1}s^{ns+1}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot(nr-1)}$$

$$R_1=\frac{nt(nt-1)\cdot\ldots\cdot(ns+3)r^{nr-2}s^{ns+2}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot(nr-2)}$$

After suitably reducing the coefficients and the pure terms by common divisors, it will be clear that M is to L_1 as $(nr+1)s$ is to $(ns)r$; L_1 is to L_2 as $(nr+2)s$ is to $(ns-1)r$; etc.; and M is to R_1 as $(ns+1)r$ is to $(nr)s$; R_1 is to R_2 as $(ns+2)r$ is to $(nr-1)s$; etc. But it is also clear that $(nr+1)s > (ns)r$; $(nr+2)s > (ns-1)r$; etc.; and that $(ns+1)r > (nr)s$; $(ns+2)r > (nr-1)s$; etc. Therefore, $M > L_1 > L_2 > \cdots$; and $M > R_1 > R_2 > \cdots$. QED.

Proof 2: It is clear that the ratio $\frac{nr+1}{ns}$ is less than the ratio $\frac{nr+2}{ns-1}$: therefore, if both ratios are multiplied by $\frac{s}{r}$, we have that

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

Similarly, it is obvious that $\frac{ns+1}{nr} < \frac{ns+2}{nr-1}$: and so, if both quantities are multiplied by $\frac{r}{s}$, we have that

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

But

$$\frac{M}{L_1}=\frac{(nr+1)\cdot s}{ns\cdot r}\quad\text{and}\quad\frac{L_1}{L_2}=\frac{(nr+2)\cdot s}{(ns-1)\cdot r};$$

likewise,

$$\frac{M}{R_1}=\frac{(ns+1)\cdot r}{nr\cdot s}\quad\text{and}\quad\frac{R_1}{R_2}=\frac{(ns+2)\cdot r}{(nr-1)\cdot s}.$$

And it can be concluded in the same way about all the other ratios in which we are interested. Therefore, the ratio between the largest term M and another term V a given distance from M is smaller than the ratio between V and another term W the same distance away from V. QED.

Lemma 4: In the expansion of a binomial whose exponent is nt, the number n can be chosen such that the ratio of M to L and the ratio of M to Λ — L and Λ being the n^{th} term away from M on the left and right respectively — will be greater than any given ratio.

Proof: Since in the preceding lemma M was found to be

$$\frac{nt(nt-1)(nt-2)\cdot\ldots\cdot(nr+1)r^{nr}s^{ns}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot ns},$$

or, equivalently

$$\frac{nt(nt-1)(nt-2)\cdot\ldots\cdot(ns+1)r^{nr}s^{ns}}{1\cdot2\cdot3\cdot4\cdot\ldots\cdot nr},$$

according to the law of expansion the terms L and Λ are as follows:

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

and

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

After making a suitable reduction by common divisors, we obtain the result that

$$\frac{M}{L} = \frac{(nr+n)(nr+n-1)(nr+n-2)\cdot \ldots \cdot (nr+1)s^n}{(ns-n+1)(ns-n+2)(ns-n+3)\cdot \ldots \cdot (ns)r^n}$$

and

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

Rewriting these two expressions after equally distributing an s (or an r) in the first n factors of the numerator and an r (or an s) in the first n factors of the denominator, and hence omitting the r^n and s^n terms, we obtain:

$$\begin{aligned}\frac{M}{L} &= \frac{\overbrace{s(nr+n)\cdot s(nr+n-1)\cdot s(nr+n-2)\cdot \ldots \cdot s(nr+1)}^{n \text{ factors}}}{\underbrace{r(ns-n+1)\cdot r(ns-n+2)\cdot r(ns-n+3)\cdot \ldots \cdot r(ns)}_{n \text{ factors}}}\\ &= \frac{(nrs+ns)(nrs+ns-s)(nrs+ns-2s)\cdot \ldots \cdot (nrs+s)}{(nrs-nr+r)(nrs-nr+2r)(nrs-nr+3r)\cdot \ldots \cdot (nrs)}\end{aligned}$$

and

$$\begin{aligned}\frac{M}{\Lambda} &= \frac{\overbrace{r(ns+n)\cdot r(ns+n-1)\cdot r(ns+n-2)\cdot \ldots \cdot r(ns+1)}^{n \text{ factors}}}{\underbrace{s(nr-n+1)\cdot s(nr-n+2)\cdot s(nr-n+3)\cdot \ldots \cdot s(nr)}_{n \text{ factors}}}\\ &= \frac{(nrs+nr)(nrs+nr-r)(nrs+nr-2r)\cdot \ldots \cdot (nrs+r)}{(nrs-ns+s)(nrs-ns+2s)(nrs-ns+3s)\cdot \ldots \cdot (nrs)}.\end{aligned}$$

But these ratios will be infinitely large when the number n is taken to be infinite; for then the numbers $1, 2, 3, \ldots$, etc. up to (but not including) n become negligible, and the numbers $nr \pm n \mp 1, 2, 3$, etc., and $ns \mp n \pm 1, 2, 3$, etc., have the same value as $nr \pm n$ and $ns \mp n$ respectively, so that by dividing through by n one has that

$$\frac{M}{L} = \frac{(rs+s)(rs+s)(rs+s)\cdot \ldots \cdot (rs)}{(rs-r)(rs-r)(rs-r)\cdot \ldots \cdot (rs)}$$

and

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

It is clear that both these quantities above are made up of the terms $\frac{rs+s}{rs-r}$ or $\frac{rs+r}{rs-s}$ as their factors. But the number of these terms is n — i.e., infinite — since the difference between the first term $nr+n$ (or $ns+n$) and the last term $nr+1$ (or $ns+1$) is $n-1$.[8] Therefore, these ratios represent $\frac{rs+s}{rs-r}$ and $\frac{rs+r}{rs-s}$ raised to an infinite power, and so are themselves infinite.

[8]Actually the number of these terms is $n-1$, for the n^{th} term in the numerator $=$ the n^{th} term in the denominator $= rs$.

If you are dubious about this conclusion, consider the infinite geometric series with the constant ratio being $\frac{rs-r}{rs+s}$ (*or* $\frac{rs-s}{rs+r}$): the ratio of the first term to the third term of this sequence is $\left(\frac{rs+s}{rs-r}\right)^2$ [*or* $\left(\frac{rs+r}{rs-s}\right)^2$]; the ratio of the first term to the fourth term is $\left(\frac{rs+s}{rs-r}\right)^3$ [*or* $\left(\frac{rs+r}{rs-s}\right)^3$]; the ratio of the first term to the fifth term is $\left(\frac{rs+s}{rs-r}\right)^4$ [*or* $\left(\frac{rs+r}{rs-s}\right)^4$]; ...; the ratio of the first term to the last term is $\left(\frac{rs+s}{rs-r}\right)^\infty$ [*or* $\left(\frac{rs+r}{rs-s}\right)^\infty$]. Moreover, it is known that the ratio of the first term to the last term is infinitely large, since the last term $= 0$. (See the corollary of our sixth treatise on Infinite Series.) Therefore, it is also known $\left(\frac{rs+s}{rs-r}\right)^\infty$ *and* $\left(\frac{rs+r}{rs-s}\right)^\infty$ are infinite. And so, it has been shown that in the expansion of a binomial raised to an infinitely high power, the greatest term M can have a ratio to L and Λ greater than any assignable ratio.

Lemma 5: In the expansion of a binomial whose exponent is nt, the number n can be chosen large enough so that the ratio of the sum of all the terms from M up to and including L (or Λ) to the sum of all the other terms beyond L (or Λ) is greater than any given ratio.

Proof: Let us use the same symbols as we used in Proof 1 of Lemma 3 above to denote the terms on either side of M; hence, in this terminology L becomes L_n and Λ becomes R_n.[9] Therefore, since by Part 2 of Lemma 3

$$\frac{M}{L_1} < \frac{L_n}{L_{n+1}} \quad \text{and} \quad \frac{L_1}{L_2} < \frac{L_{n+1}}{L_{n+2}} \quad \text{and} \quad \frac{L_2}{L_3} < \frac{L_{n+2}}{L_{n+3}} \quad \text{and} \quad \ldots,$$

we also have that

$$\frac{M}{L_n} < \frac{L_1}{L_{n+1}} < \frac{L_2}{L_{n+2}} < \frac{L_3}{L_{n+3}} \quad \ldots.$$

Thus, when n is taken to be infinite, the ratio $\frac{M}{L_n}$ will be infinitely large and, by Lemma 4, the other ratios $\frac{L_1}{L_{n+1}}$, $\frac{L_2}{L_{n+2}}$, $\frac{L_3}{L_{n+3}}$, ... will be all the more infinite, and as a result

$$\frac{L_1 + L_2 + L_3 + \ldots + L_n}{L_{n+1} + L_{n+2} + L_{n+3} + \ldots + L_{2n}}$$

will be infinite: i.e., all the terms summed from L_1 to L_n inclusive will be infinitely greater than the terms summed from L_{n+1} to L_{2n} inclusive. And since, by Lemma 1, the number of all the terms beyond L_n[10] is only $(s-1)$ times — that is, a finite number of times — the number of terms between M and L_n,[11] and since by Part 1 of Lemma 3 the terms become smaller the farther they are from L_n, all the terms added together from M (but not including M) down to and including L_n will still be infinitely greater than all the terms added together beyond L_n. Moreover, it is shown in the same way for the other direction that all the terms between R_1 and R_n inclusive are infinitely larger than all the terms beyond R_n. (Note that the number of terms beyond R_n is, by Lemma 1, only $(r-1)$ times

[9] For this terminology — which has substantially clarified the presentation of the following mathematical proofs — I am indebted to Herr R. Haussner, whose 1899 German translation of the *Ars Conjectandi* has been of considerable use to me. (In Bernoulli's original notation: $L_1 = F$; $L_2 = G$; $L_3 = H$; etc.; $L_{n+1} = P$; $L_{n+2} = Q$; $L_{n+3} = R$; etc. The advantages of Haussner's notation are, I think, quite clear.)

[10] Here, L_k is "beyond" L_j if and only if $k < j$; likewise, R_k is "beyond" R_j if and only if $k > j$.

[11] Actually, Lemma 1 says that the number of terms beyond L_n is $(s-1)$ times the number of terms between M down to and *including* L_n.

the number of terms between M and R_n.[12]) Hence, all the terms added together between L_n and R_n inclusive except M will likewise infinitely exceed all the terms added together beyond these limits L_n and R_n. Therefore, all the terms summed together between L_n and R_n inclusive — i.e., the former sum plus M — will all the more exceed all the terms summed together beyond L_n and R_n. **QED**.

Remark: Those who have not been accustomed to deal with speculations about infinity may say the following in objection to Lemmas 4 and 5: although in the case of n being an infinite number, the factors of the quantities representing $\frac{M}{L_n}$ and $\frac{M}{R_n}$ — i.e., $nr \pm n \mp 1, 2, 3, \ldots$ and $ns \mp n \pm 1, 2, 3, \ldots$ — have the same value as $nr \pm n$ and $ns \mp n$ (since the numbers $1, 2, 3, \ldots$ become negligible), it can nevertheless happen that when all these terms are taken or multiplied together, they will yield an infinite number (since the number of factors is infinite); and so, the infinitely high power of the quotient $\frac{rs+s}{rs-r}$ or $\frac{rs+r}{rs-s}$ will infinitely diminish — that is, it will become finite. I cannot better satisfy this doubt than if I now show a way of ascertaining that for any truly finite number n or for any finite power of the binomial, the *sum* of the terms between L_n and R_n inclusive *has a ratio to the sum* of the terms beyond L_n and R_n *greater than any given ratio*, say c. Having proven this fact, I will have shown that the objection must of its own accord be destroyed.

Consider the terms to the left of M: To this end, I take any given ratio greater than 1 but less than $\frac{rs+s}{rs-r}$, say $\frac{rs+s}{rs} = \frac{r+1}{r}$, and I multiply it by itself as often — m times — as necessary for it to be greater than or equal to $c(s-1)$: i.e., so that

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

Moreover, it is shortly found by logarithms what the value of m should be; for after the logarithms of the quantities have been taken, we have that

$$m\log(r+1) - m\log(r) \geq \log[c(s-1)],$$

and after performing a division we immediately find that

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

Having found this, I proceed as follows: In that series of fractions or factors

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

— the product of which, by Lemma 4, results in the quotient $\frac{\mathrm{M}}{\mathrm{L}_n}$ — one may observe that each of the fractions is less than $\frac{rs+s}{rs-r}$, although they come continually nearer to it the larger the value of n; and so, one of these terms must at sometime be equal to $\frac{rs+s}{rs} = \frac{r+1}{r}$.

Therefore, one should see that the value n must have is such that the fraction which occupies the m^{th} position of the above series equals $\frac{r+1}{r}$. But, as is clear from the formulation of the series, the fraction in the m^{th} position is $\frac{nrs+ns-ms+s}{mrs-nr+mr}$, which, when equated to $\frac{r+1}{r}$, gives $n = m + \frac{ms-s}{r+1}$, and thence $nt = mt + \frac{mst-st}{r+1}$. I claim that this will be the exponent to which the binomial $(r+s)$ should be raised for the largest term M to exceed the bound L_n by more than $c(s-1)$ times. For since — by this assumption of the number n — the fraction in the m^{th} position is $\frac{r+1}{r}$, and since by hypothesis

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

[12]Similarly, the number of terms beyond R_n is actually $(r-1)$ times the number of terms between M down to and *including* R_n.

we have that this fraction $\frac{r+1}{r}$ multiplied by all the preceding ones exceeds $c(s-1)$ by much more, for each of the preceding fractions is greater than $\frac{r+1}{r}$; also, since each of the terms beyond the m^{th} term is at least greater than 1, the product of all the terms in this series will exceed $c(s-1)$ by still more. But the product of all these fractions is $\frac{M}{L_n}$; therefore, it is agreed by all that M exceeds the bound L_n by more than $c(s-1)$ times.

Now, moreover,

$$\frac{M}{L_n} < \frac{L_1}{L_{n+1}} < \frac{L_2}{L_{n+2}} < \frac{L_3}{L_{n+3}} \quad \ldots,$$

as has been shown. Hence, L_1 will exceed L_{n+1} by much more than $c(s-1)$ times, and L_2 will exceed L_{n+2} by even more, etc. And so, finally, all the terms beyond M down to and including L_n exceed by more than $c(s-1)$ times the same number of terms extending beyond L_n — i.e., the terms beyond L_n down to and including L_{2n}; thus, the former sum of terms exceeds $(s-1)$ times the latter sum of terms by more than c times. Therefore, much more clearly do the terms beyond M down to and including L_n exceed all the terms beyond L_n by more than c times, since the entire latter group of terms outnumber the entire former group of terms by only $(s-1)$ times.

Consider the terms to the right of M: I proceed in similar fashion: I take the ratio $\frac{s+1}{s} < \frac{rs+r}{rs-s}$ and I make

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

I then find that

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

Next, in the series of fractions

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

(all of which make up the quotient $\frac{M}{R_n}$), I assume that the fraction in the m^{th} position — i.e., $\frac{nrs+nr-mr+r}{nrs-ns+ms}$ — equals $\frac{s+1}{s}$, and thence I deduce that

$$n = m + \frac{mr-r}{s+1},$$

and that therefore

$$nt = mt + \frac{mrt-rt}{s+1}.$$

Having done this, I show similarly as before that when the binomial $(r+s)$ has been raised to this power nt, the largest term M will exceed the bound R_n by more than $c(r-1)$ times; and as a consequence also, that all the terms beyond M up to and including R_n will exceed all those terms beyond R_n by more than c times. (Note that the latter group of terms outnumber the former group by only $(r-1)$ times.) And so at last we finally conclude that when the binomial $(r+s)$ is raised to that power whose numerical value is equal to the larger of the two quantities $mt + \frac{mst-st}{r+1}$ and $mt + \frac{mrt-rt}{s+1}$, all the terms added together between L_n and R_n inclusive will exceed by more than c times all the terms added together beyond L_n and beyond R_n. Therefore, a finite power has been found which has the desired property. **QEF**.

Main Theorem: Finally, *the* theorem follows upon which all that has been said is based, but whose proof now is given solely by the application of the lemmas stated above. In order that I may avoid being tedious, I will call those cases in which a certain event can happen *successful* or *fertile* cases; and those cases sterile in which the same event cannot happen. Also, I will call those trials *successful* or *fertile* in which any of the fertile cases is perceived; and those trials *unsuccessful* or *sterile* in which any of the sterile cases is observed. Therefore, let the number of fertile cases to the number of sterile cases be exactly or approximately in the ratio r to s, and hence the ratio of fertile cases to all the cases will be $\frac{r}{r+s}$ or $\frac{r}{t}$, which is within the limits $\frac{r+1}{t}$ and $\frac{r-1}{t}$. It must be shown that so many trials can be run such that it will be more probable than any given times (e.g., c times) that the number of fertile observations will fall within these limits rather than outside these limits — i.e., it will be c times more likely than not that the number of fertile observations to the number of all the observations will be in a ratio neither greater than $\frac{r+1}{t}$ nor less than $\frac{r-1}{t}$.

Proof: Let nt be the number of observations taken, and let it be asked what is the expectation or what is the probability that all are successful with none being sterile; with one being sterile; with two being sterile; with three being sterile; with four being sterile; etc. Since, moreover, in any observation there are at hand, by hypothesis, t cases of which r are successful and s are sterile, and since each case of one observation can be combined with each case of a second observation, and can again be combined with each case of a third observation, and with each case of a fourth observation, etc., it is clearly evident that the rule tied in at the end of the notes to Prop. XIII of the first Part and the second corollary of this rule are applicable to this task. The second corollary contains the general formula, and with its help it is known that the expectation of no observations being sterile is $\frac{r^{nt}}{t^{nt}}$; of one observation being sterile

$$\frac{\frac{nt}{1}\cdot r^{nt-1}\cdot s^1}{t^{nt}};$$

of two observations being sterile,

$$\frac{\frac{nt(nt-1)}{1\cdot 2}\cdot r^{nt-2}\cdot s^2}{t^{nt}};$$

of three observations being sterile,

$$\frac{\frac{nt(nt-1)(nt-2)}{1\cdot 2\cdot 3}\cdot r^{nt-3}\cdot s^3}{t^{nt}};$$

and so on. Thus, after the common denominator t^{nt} has been omitted, it is clear that the degrees of the probabilities or the numbers of cases in which it can happen that all the trials are successful, or that all except one, or two, or three, or four, etc., are successful may be expressed in order by

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

But these are the very terms of the expansion of the binomial $(r+s)$ raised to the $(nt)^{\text{th}}$ power, an expansion examined above in our lemmas; whence all the rest is quite evident:

For according to the nature of the expansion, it is clear that the number of cases which result in ns sterile trials and nr successful ones is the largest term M of the expansion, in as much as ns terms precede M and nr terms follow M according to Lemma 3. Likewise, it is clear that the numbers of those cases in which either $nr+n$ or $nr-n$ trials are successful — the rest of the trials being sterile — are expressed by the terms of the expansion L_n and R_n respectively, where L_n and R_n are defined as in the proof of Lemma 5 above. Also as a consequence, it is clear that the sum of the cases in which not more than $nr+n$ trials nor less than $nr-n$ trials are successful may be expressed by the sum of the terms in the expansion between L_n and R_n inclusive; the sum of the other cases in which either more than $nr + n$ or less than $nr - n$ successful trials are realized is expressed by the sum of the terms outside these bounds L_n and R_n.

Let $R =$ the ratio of the number of fertile observations to the number of all observations. By Lemmas 4 and 5, a power of the binomial can be chosen to be so large that the sum of the terms between L_n and R_n inclusive exceeds the sum of the terms outside these bounds by more than c times; therefore, it follows that so many observations can be made such that the sum of the cases in which R lies between $\frac{nr+n}{nt}$ and $\frac{nr-n}{nt}$ (or, equivalently, between $\frac{r+1}{t}$ and $\frac{r-1}{t}$) exceeds the sum of the other cases by more than c times; that is, it is more than c times *more* likely that R will fall between the limits $\frac{r+1}{t}$ and $\frac{r-1}{t}$ *than* that R will fall outside these limits. **QED**.

Moreover, in the specific application of these things to numbers, it is clear enough that the larger the numbers r, s, and t are taken to be (always remaining in the same ratio to each other), the tighter the limits $\frac{r+1}{t}$ and $\frac{r-1}{t}$ can be drawn around the ratio $\frac{r}{t}$. On this account, for example, if the ratio $\frac{r}{s}$ (which must be determined by trials) is $3/2$, I do not take $r = 3$ and $s = 2$, bur rather $r = 30$ and $s = 20$, or $r = 300$ and $s = 200$, etc. It would be sufficient to take $r = 30$ and $s = 20$ (and hence $t = r + s = 50$) for the limits to be $\frac{r+1}{t} = \frac{31}{50}$ and $\frac{r-1}{t} = \frac{29}{50}$; further, let $c = 1000$. It would thus follow from the remarks above that for the terms on the left:

$$m > \frac{\log[c(r-1)]}{\log(s+1) - \log(s)} = \frac{4.2787536}{.0142405} < 301,$$

and

$$nt = mt + \frac{mst - st}{r+1} < 24,728;$$

and for the terms on the right:

$$m > \frac{\log[c(s-1)]}{\log(r+1) - \log(r)} = \frac{4.4623980}{.0211893} < 211,$$

and

$$nt = mt + \frac{mrt - rt}{s+1} < 25,550.$$

Whence, by what has already been proven, it is concluded that in 25,550 trials it is more than one thousand times more likely that $\frac{r}{t}$ will fall between $\frac{31}{50}$ and $\frac{29}{50}$ than that $\frac{r}{t}$ will fall outside these limits. (Note that $\frac{r}{t}$ is the ratio of the number of fertile observations to the number of all observations.) And in the same way, if c is taken to be 10,000, one will find that 31,258 trials are necessary for it to be 10,000 times more likely that $\frac{r}{t}$ will fall between $\frac{31}{50}$ and $\frac{29}{50}$ than that $\frac{r}{t}$ will fall outside these limits; and if c is taken to be 100,000, one will find that 36,966 trials are necessary for it to be 100,000 times more likely that $\frac{r}{t}$ will fall between $\frac{31}{50}$ and $\frac{29}{50}$ than that $\frac{r}{t}$ will fall outside these limits; etc.; and so on *ad infinitum*, always adding a multiple of 5,708 trials to 25,550. Whence, finally, this one thing seems to follow: that if observations of all events were to be continued throughout all eternity, (and hence the ultimate probability would tend toward perfect certainty), everything in the world would be perceived to happen in fixed ratios and according to a constant law of alternation, so that even in the most accidental and fortuitous occurrences we would be bound to recognize, as it were, a certain necessity and, so to speak, a certain fate.

I do not know whether Plato wished to aim at this in his doctrine of the universal return of things, according to which he predicted that all things will return to their original state after countless ages have past.

CORRESPONDENCE BETWEEN LEIBNITZ AND BERNOULLI

The following excerpts are translated from Leibnitz (1855).

P. 71. Short Paragraph: "P.S. . . . omnes." (Leibnitz in a post script of a letter to James Bernoulli, April, 1703.)

"**P.S.**: I hear that the subject of estimating probabilities — which I consider important — has been not a little developed by you. I would like someone to treat mathematically the various kinds of games (in which there are beautiful examples of this subject.) This task would be both pleasant and useful, and it would not be unworthy of you or any very serious mathematician. I have seen some of your stated theses and only a few discussions of them. However, I would like to have them all.

P. 77–8. Long paragraph: "Scire libenter. . . fieri percuperem." (An excerpt from a letter to Leibnitz from James Bernoulli, 3 October, 1703.)

"I would gladly like to know, most honorable sir, from whom you know that I have been working on the subject of estimating probabilities. It is true that for many years past I have taken much pleasure in explorations of this sort, since I scarcely think that anyone else has thought more than I about these matters. I even had a mind to write a tract about this subject; but I have often put it off for years at a time, because my natural laziness — which the weakness of my health, as an accomplice, has increased so much more — caused me to approach the writing very feebly. I often wish I had a secretary who could fully divine my thoughts when they were gently hinted to him, and could put them down in writing. Nevertheless, I have already completed the larger part of a book, but with an important part still missing, in which I show how to apply the principles of the art of estimation to civil, moral, and economic affairs. I will finish the book after I have solved a singular problem, which has a not small commendation of difficulty and a very large commendation of usefulness, and which has remained before my brother for twelve years, although he, when asked about the same problem some time ago by Marquis de l'Hospital, concealed the truth because of his eagerness to devalue my research. I will briefly tell you what the problem is: it is a known fact that the probability of any event depends on the number of possible outcomes with which the event can or cannot happen; and so, it occurred to me to ask why, for example, do we know with how much greater probability a seven rather than an eight will fall when we roll a pair of dice, and why indeed do we *not* know how much more probable it is for a young man of twenty years to survive an old man of sixty years than for an old man of sixty years to survive a young man of twenty years; this is the point: we know the number of possible ways in which a seven and in which an eight can fall when rolling dice, but we do not know the number of possible ways which prevail in summoning a young man to die before an old man, and which prevail in summoning an old man to die before a young man. I began to inquire whether what is hidden from us by chance *a priori* can at least be known by us *a posteriori* from an occurrence observed many times in similar cases — i.e., from an experiment performed on many pairs of young and old men. For had I observed it to have happened that a young man outlived his respective old man in one thousand cases, for example, and to have happened otherwise only five hundred times, I could safely enough conclude that it is twice as probable that a young man outlives an old man as it is that the latter outlives the former. Moreover, although — and this is amazing — even the stupidest man knows, by some instinct of nature *per se* and by no previous instruction, that the more observations there are, the less danger there is in straying from the mark, it requires not at all ordinary research to demonstrate this fact accurately and geometrically. But this is not all that I want: in addition, it must be inquired whether the probability of an accurate ratio increases steadily as the number of observations grows, so that finally the probability that I have found the true ratio rather than a false ratio exceeds any given probability; or whether each problem, so to speak, has an asymptote — that is, whether I shall finally reach some level of probability beyond which I cannot be more certain that I have detected the true ratio. For if the latter is true, we will be done with our attempt at finding out the number of possible outcomes through experiments; if the former is true, we will investigate the ratio between the numbers of possible outcomes *a posteriori* with as much certainty as if it were known to us *a priori*. And I have found the former condition is indeed the case; whence I can now determine how many trials must be set up so that it will be a hundred, a thousand, ten thousand, etc., times more probable (and finally, so that it will be morally certain) that the ratio between the numbers of possible outcomes which I obtain in this way is legitimate and genuine. The following suffices for practice in civil life: to formulate our conjectures in any situation that may occur no less scientifically than in game of chance; I think that all the wisdom of a politician lies in this alone. I do not know, most honorable sir, whether anything of substance appears to you to be in these speculations; in any case, you will make me grateful if you could supply me with any legal situations which you think could be usefully applied to these matters. Recently, I found that a certain tract which had been unknown to me was cited in the printed Monthly Excerpts of Hanover: Pensionarius de Wit's *von Subtiler Ausrechnung des valoris der Leib-Renten*. Perhaps he has something doing here; whatever it is, I would very much wish to obtain his source from somewhere."

P. 83-4. Long paragraph: "Utilissima est... apparet." (An excerpt from a letter to James Bernoulli from Leibnitz, 3 December, 1703.)

"The estimation of probabilities is extremely useful, although in several political and legal situations there is not much need for fine calculation as there is for the accurate recapitulation of all the circumstances. I remember learning for the first time *not* from your brother but from somewhere else that these matters had been dealt with by you. When we estimate empirically, by means of experiments, the probabilities of successes, you ask whether a perfect estimation can be finally obtained in this manner. You write that you have found this to be so. There appears to me to be a difficulty in this conclusion: that happenings which depend upon an infinite number of cases cannot be determined by a finite number of experiments; indeed, nature has her own habits, born from the return of causes, but only 'in general.' And so, who will say whether a subsequent experiment will not stray somewhat from the rule of all the preceding experiments, because of the very mutabilities of things? New diseases continually inundate the human race, but if you had performed as many experiments as you please on the nature of deaths, you have not on that account set up the boundaries of the world so that it cannot change in the future. When we investigate the path of a comet from any number of observations, we suppose that it is either a conic curve or another kind of simple curve. Given any number of points, an infinite number of curves can be found passing through them. Thus, I show the following: I postulate (and this can be demonstrated) that given any number of points, some regular curve can be found passing through these points. Let it be given that this curve has been found, and call it "A. Now, let another point be taken lying between the points given but outside of this curve; let a curve pass through this new point and the points given originally, according to the above postulate: this curve must be different from the first curve, but nevertheless it passes through the same given points through which the first curve passes. And since a point can be varied an infinite number of times, there will also be an infinite number of these and other possible curves. Moreover, observed outcomes can be compared with these points, where the fixed underlying outcomes or their estimates inferred from observed outcomes, can be compared with the model curve. It may be added that, although a perfect estimation cannot be had empirically, an empirical estimate would nonetheless be useful and sufficient in practice. The person who composed the monthly Germanic excerpts of Hanover has been at my house. Pensionarius de Wit's article is flimsy when he uses that estimation known from the equal possibility of similar outcomes and hence shows that the problem of resurrections can be clearly solved by considering the fate of the Batavians.[13] And therefore, he has written in Belgic, so that he might appear to be on the same footing with the commoner.

P. 87-9. Long Paragraph: "Quod Doctrina... ventilata fuit." (Selection from a letter to Leibnitz from James Bernoulli, 20 April, 1704.)

"Various questions about Certainty, Resurrections, Endowed Agreements, Conjectures, and other matters show me that the subject of estimating probabilities in legal affairs requires not only the recapitulation of circumstances but also the same computation and calculation which we are accustomed to use in reckoning the outcomes of games of chance; I will show how to do this clearly for each situation. Moreover, the difficulty you found with my empirical method in determining the ratio between the numbers of possible outcomes requires more examples, not those in which it is impossible by any means to agree upon the numbers themselves, but rather those in which the numbers can be learned *a priori*. In addition, I said that I could, in these examples, provide for you a demonstration (which my brother saw twelve years ago and approved of). In order that you may really understand more clearly what I think, I give you an example: I place in an urn several hidden pebbles, black and white ones, and the number of white ones is twice the number of black ones; but you do not know this ratio, and you wish to determine it by experiment. And so, you draw one pebble out after another (replacing the pebble which you drew out in each single choice before you draw the next one, so that the number of pebbles in the urn is not diminished,) and you note whether you have picked a white or a black one. Now, I claim (assuming that you have two estimates of the two-to-one ratio which are, though quite close to one another, different, one being larger, the other being smaller — say $201 : 100$ and $199 : 100$) that I can determine scientifically the necessary number of observations so that with ten, a hundred, a thousand, etc. times more probability, the ratio of the number of drawings in which you choose a white pebble to the number of drawings in which you choose a black pebble will fall within, rather than outside of, these limits of the two-to-one ratio: $201 : 100$ and $199 : 100$; and so I claim that you can be morally certain the ratio obtained by experiment will come as close as you please to the true two-to-one ratio. But if now in place of the urn you substitute the human body of an old man or a young man, the human body which contains the tinder of sicknesses within itself as the urn contains pebbles, you can determine in the same way through observations how much nearer to death the one is than the other. It does no good to say that the number of sicknesses to which each is exposed is infinite; for let us grant this; it is nevertheless known that there are levels in infinity, and that the ratio of one infinity to another infinity is still a finite number, and can be expressed either precisely or sufficiently precisely for practical use. If sicknesses are multiplied by

[13]My interpretation of this is the following:
Problem: What is the possibility of resurrection?
Solution: Look at the proportion of Batavians who have been resurrected. (De Wit himself was a Batavian.)

the passage of time, then new observations, in any case, must be set up: and it is certain that he who thinks that the investigations of our ancient forefathers concerning the end of life be settled by the daily customary observations made in London, Paris, or elsewhere will grossly err from the truth. The example of investigating the trajectory of a comet from several of its observed positions is, in this situation, almost apropos; I would never use it to demonstrate a proposition: although, in a limited way, I can find an application, since it cannot be denied that if five points have been observed, all of which are perceived to lie along a parabola, the notion of a parabola will be stronger than if only four points had been observed: for although there are an infinite number of curves which may pass through five points, there is nevertheless beyond this infinite number another infinite number — rather, an infinitely times more infinite number — of curves which may pass through only the first four points and not through the fifth point, all of which are excluded by this fifth observation. And yet, I admit that every conjecture which is deduced by observations of this sort would be quite flimsy and uncertain if it were not conceded that the curve sought is one of the class of simple curves; this indeed seems quite correct to me, since we see everywhere that nature follows the simplest paths. I perceive from your description that the Belgian tract of Jean de Wit contains such things which serve my point very well. And so I ask as strongly as possible that you, most honorable sir, send to me your copy of the book on any convenient occasion, since I have sought for it in vain in Amsterdam. I shall return it faithfully on the next market day in Frankfurt together with the fourth and fifth part of my publications concerning infinite series, the latter of which has been recently published and circulated."

ACKNOWLEDGMENT

I want to express my deep gratitude to Dr. Alan Gillingham of Phillips Academy, Andover, for his constant encouragement while I was doing this translation and for the counsel and assistance he gave me in construing some of the more difficult passages.

Bing Sung

February, 1966

REFERENCES

BERNOULLI, JAMES (1713) *Ars Conjectandi*

DAVID, F.N. (1962) *Games, Gods and Gambling*. Hafner: New York.

HAUSSNER, R. (1899) *Wahrscheinlichkeitsrechnung* (A German translation of Parts Three and Four of Ars Conjectandi together with a letter to a friend about the game of ball.) Engelmann: Leipsig.

LEIBNITZ (1855) *Leibnizens Mathematische Schriften herausgegeben von C.I. Gerhardt*. Erste Abteilung.

PEARSON, KARL (1925) James Bernoulli's Theorem. *Biometrika*, 17, 201–10.

TODHUNTER, F. (1865) *A History of the Mathematical Theory of Probability*.