Baye's "An Essay towards solving a Problem in the Doctrine of Chances" and Notes on Bayes' Essay

Shankar Raman

Since Bayes' groundbreaking contribution to probability is not easy to read, I have adopted the tactic of excerpting the sections of relevance to us, and interspersing a commentary that explicates some of the key ideas. The sections you are expected to read are the following:

1. Price's letter introducing Bayes' essay.
2. Section I, consisting of the Definitions and Proposition 1 - 7: Proposition 4 is very difficult to reconstruct, so just focus on the result.
3. In Section 2 focus on the laying out of the problem, the results of Postulate 1 and 2, the result of Proposition 8, Proposition 9 and the Scholium. Don't worry about the demonstrations in this section, since my commentary reconstructs the argument being made in a way that is easier to follow.

Before turning to the essay itself, however, it may be useful to set out some basic ideas of probability theory in its modern form to aid the understanding of Bayes' thought experiment.

Basic Ideas:

Conditional Probability: Probabilities are always relative to a universe, which we have called Ω. If we think, however, of a set of events in that universe in terms of what it shares with a different *subset* of that universe, the latter in effect defines a new universe. And this is what conditional probability is about: looking at events as conditioned upon or given a subset of Ω as domain of concern. The conditional probability of a set A given a set B (both subsets of Ω) is written as $P(A|B)$. This conditional probability can be expressed in terms of the unconditional (that is, relative to Ω) probabilities of A and B. Bayes will attempt below to "prove" that relationship. But modern probability theory simply defines unconditional probability as this relationship (thereby creating a new conditional space for which the basic axioms of probability are valid). Let me sum up the definitions:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(B) * P(A|B) = P(A \cap B) \tag{1}$$

and likewise

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A) * P(B|A) = P(A \cap B) \tag{2}$$

Intuitively, what these equations express is that a set of events shared by two subsets A and B in the universe – in other words, the events comprising $A \cap B$ – can be described from different perspectives: (1) We can look at them unconditionally as a subset of the universe as a whole (in which case their probability is written as $P[A \cap B]$). Or (2) we can view those events as events in B conditioned upon A's happening and ask about this subset of B type events relative to A (i.e $P[B|A]$). Or (3) we can view these as events in A conditioned upon B's happening (i.e., $P[A|B]$).

Independence: Two sets of events A and B are independent if the occurrence of one gives us no information regarding the occurrence of the other. Or, A's happening has no effect on B's happening (and vice versa). This relationship can be stated as a relationship of conditional and unconditional probabilities. Viz. $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Alternatively, using the equation 1 for $P(A|B)$ above and rearranging terms: if A and B are independent, $P(A \cap B) = P(A) * P(B)$.

NOTE: Independence is NOT THE SAME as disjunction. Two events are disjunct if $A \cap B = \emptyset$, that is, A and B share no events in common. But independence *requires* that they share events in common. Why? Well, if A and B are disjunct, this means that if A happens, B does not happen and vice versa. Therefore the happening of A gives us a great deal of information about B because it tells us that we can be certain that B did not happen. But independence requires that the happening of A has no bearing on the happening of B.

Let's see this via an example. If I toss a coin three times in succession, each toss is clearly independent of the other – getting a head on the first tells me nothing about what I may get in on the second or the third. For each toss, the outcome space is the same, that is, $\Omega = \{H, T\}$, and, for instance, P(toss 1 = head, given that toss 2 = head) = P (toss 1 = head). Now consider, the set of outcomes of 3 tosses, That is, $\Omega = \{HHH, HHT, HTH, HTT, TTT, TTH, THT, THH\}$. These outcomes are mutually exclusive, that is, they are disjunct, so the happening of any one tells us that the other cannot happen. Consequently, they are not independent.

The Multiplication Rule: This rule allows us to write the joint probability of a series of events in terms of their conditional probabilities. Say we have 3 events A, B and C, then the probability that all three will occur (that is, $P[A \cap B \cap C]$) can be calculated by multiplying the unconditional probability of A happening, by the conditional probability that B happens given that A happened, by the conditional probability that C happens given that both A and B happened. That is,

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

This can be simply verified by plugging in the appropriate terms for the conditional probabilities using equation 1 above. And if you draw this out as a tree or think about the Venn diagram, you will get an intuitive sense of this result. And the multiplication rule can be extended to any number of events. A, B and C need not be independent. If they are independent, the equation gets simpler, since all the conditionals disappear, leaving

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

In other words, if events are independent the likelihood of all of them happening is simply the product of the (unconditional) individual probabilities of their occurrence.

AN ESSAY TOWARDS SOLVING A PROBLEM IN THE DOCTRINE OF CHANCES

BY THE LATE REV. MR BAYES, F.R.S.

Communicated by Mr Price, in a Letter to John Canton, A.M., F.R.S.

Read 23 *December* 1763

Dear Sir,

I now send you an essay which I have found among the papers of our deceased friend Mr Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illustrious Society, and was much esteemed by many in it as a very able mathematician. In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times. He adds, that he soon perceived that it would not be very difficult to do this, provided some rule could be found according to which we ought to estimate the chance that the probability for the happening of an event perfectly unknown, should lie between any two named degrees of probability, antecedently to any experiments made about it; and that it appeared to him that the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees; which, if it were allowed, all the rest might be easily calculated in the common method of proceeding in the doctrine of chances. Accordingly, I find among his papers a very ingenious solution of this problem in this way. But he afterwards considered, that the *postulate* on which he had argued might not perhaps be looked upon by all as reasonable; and therefore he chose to lay down in another form the proposition in which he thought the solution of the problem is contained, and in a *scholium* to subjoin the reasons why he thought so, rather than to take into his mathematical reasoning any thing that might admit dispute. This, you will observe, is the method which he has pursued in this essay.

Every judicious person will be sensible that the problem now mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to [provide] a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter. Common sense is indeed sufficient to shew us that, from the observation of what has in former instances been the consequence of a certain cause or action, one may make a judgment what is likely to be the consequence of it another time, and that the larger [the] number of experiments we have to support a conclusion, so much the more reason we have to take it for granted. But it is certain that we cannot determine, at least not to any nicety, in what degree repeated experiments confirm a conclusion, without the particular discussion of the beforementioned problem; which, therefore, is necessary to be considered by any one who would give a clear account of the strength of *analogical* or *inductive reasoning*; concerning, which at present, we seem to know little more than that it does sometimes in fact convince us, and at other times not; and that, as it is the means of [a]cquainting us with many truths, of which otherwise we must have been ignorant; so it is, in all probability, the source of many errors, which perhaps might in some measure be avoided, if the force that this sort of reasoning ought to have with us were more distinctly and clearly understood.

These observations prove that the problem enquired after in this essay is no less important than it is curious. It may be safely added, I fancy, that it is also a problem that has never before been solved. Mr De Moivre, indeed, the great improver of this part of mathematics, has in his *Laws of Chance*,* after Bernoulli, and to a greater degree of exactness, given rules to find the probability there is, that if a very great number of trials be made concerning any event, the proportion of the number of times it will happen, to the number of times it will fail in those trials, should differ less than by small assigned limits from the proportion of the probability of its happening to the probability of its failing in one single trial. But I know of no person who has shewn how to deduce the solution of the converse problem to this; namely, 'the number of times an unknown event has happened and failed being given, to find the chance that the probability of its happening should lie somewhere between any two named degrees of probability.' What Mr De Moivre has done therefore cannot be thought sufficient to make the consideration of this point unnecessary: especially, as the rules he has given are not pretended to be rigorously exact, except on supposition that the number of trials made are infinite; from whence it is not obvious how large the number of trials must be in order to make them exact enough to be depended on in practice.

Mr De Moivre calls the problem he has thus solved, the hardest that can be proposed on the subject of chance. His solution he has applied to a very important purpose, and thereby shewn that those are much mistaken who have insinuated that the Doctrine of Chances in mathematics is of trivial consequence, and cannot have a place in any serious enquiry.† The purpose I mean is, to shew what reason we have for believing that there are in the constitution of things fixt laws according to which events happen, and that, therefore, the frame of the world must be the effect of the wisdom and power of an intelligent cause; and thus to confirm the argument taken from final causes for the existence of the Deity. It will be easy to see that the converse problem solved in this essay is more directly applicable to this purpose; for it shews us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any of the irregularities of chance.

The two last rules in this essay are given without the deductions of them. I have chosen to do this because these deductions, taking up a good deal of room, would swell the essay too much; and also because these rules, though of considerable use, do not answer the purpose for which they are given as perfectly as could be wished. They are however ready to be produced, if a communication of them should be thought proper. I have in some places writ short notes, and to the whole I have added an application of the rules in the essay to some

* See Mr De Moivre's *Doctrine of Chances*, p. 243, etc. He has omitted the demonstrations of his rules, but these have been since supplied by Mr Simpson at the conclusion of his treatise on *The Nature and Laws of Chance*.

† See his *Doctrine of Chances*, p. 252, etc.

particular cases, in order to convey a clearer idea of the nature of the problem, and to shew how far the solution of it has been carried.

I am sensible that your time is so much taken up that I cannot reasonably expect that you should minutely examine every part of what I now send you. Some of the calculations, particularly in the Appendix, no one can make without a good deal of labour. I have taken so much care about them, that I believe there can be no material error in any of them; but should there be any such errors, I am the only person who ought to be considered as answerable for them.

Mr Bayes has thought fit to begin his work with a brief demonstration of the general laws of chance. His reason for doing this, as he says in his introduction, was not merely that his reader might not have the trouble of searching elsewhere for the principles on which he has argued, but because he did not know whither to refer him for a clear demonstration of them. He has also made an apology for the peculiar definition he has given of the word *chance* or *probability*. His design herein was to cut off all dispute about the meaning of the word, which in common language is used in different senses by persons of different opinions, and according as it is applied to *past* or *future* facts. But whatever different senses it may have, all (he observes) will allow that an expectation depending on the truth of any *past* fact, or the happening of any *future* event, ought to be estimated so much the more valuable as the fact is more likely to be true, or the event more likely to happen. Instead therefore, of the proper sense of the word *probability*, he has given that which all will allow to be its proper measure in every case where the word is used. But it is time to conclude this letter. Experimental philosophy is indebted to you for several discoveries and improvements; and, therefore, I cannot help thinking that there is a peculiar propriety in directing to you the following essay and appendix. That your enquiries may be rewarded with many further successes, and that you may enjoy every valuable blessing, is the sincere wish of, Sir,

your very humble servant,

Richard Price

Newington-Green,
10 *November* 1763

PROBLEM

Given the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

Section I

Definition 1. Several events are *inconsistent*, when if one of them happens, none of the rest can.

2. Two events are *contrary* when one, or other of them must; and both together cannot happen.

3. An event is said to *fail*, when it cannot happen; or, which comes to the same thing, when its contrary has happened.

4. An event is said to be determined when it has either happened or failed.

5. The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon it's happening.

6. By *chance* I mean the same as probability.

7. Events are independent when the happening of any one of them does neither increase nor abate the probability of the rest.

Prop. 1

When several events are inconsistent the probability of the happening of one or other of them is the sum of the probabilities of each of them.

Suppose there be three such events, and whichever of them happens I am to receive N, and that the probability of the 1st, 2nd, and 3rd are respectively a/N, b/N, c/N. Then (by the definition of probability) the value of my expectation from the 1st will be a, from the 2nd b, and from the 3rd c. Wherefore the value of my expectations from all three will be $a+b+c$. But the sum of my expectations from all three is in this case an expectation of receiving N upon the happening of one or other of them. Wherefore (by definition 5) the probability of one or other of them is $(a+b+c)/N$ or $a/N+b/N+c/N$. The sum of the probabilities of each of them.

Corollary. If it be certain that one or other of the three events must happen, then $a+b+c = N$. For in this case all the expectations together amounting to a certain expectation of receiving N, their values together must be equal to N. And from hence it is plain that the probability of an event added to the probability of its failure (or of its contrary) is the ratio of equality. For these are two inconsistent events, one of which necessarily happens. Wherefore if the probability of an event is P/N that of it's failure will be $(N-P)/N$.

Prop. 2

If a person has an expectation depending on the happening of an event, the probability of the event is to the probability of its failure as his loss if it fails to his gain if it happens.

Suppose a person has an expectation of receiving N, depending on an event the probability of which is P/N. Then (by definition 5) the value of his expectation is P, and therefore if the event fail, he loses that which in value is P; and if it happens he receives N, but his expectation ceases. His gain therefore is $N-P$. Likewise since the probability of the event is P/N, that of its failure (by corollary prop. 1) is $(N-P)/N$. But P/N is to $(N-P)/N$ as P is to $N-P$, i.e. the probability of the event is to the probability of it's failure, as his loss if it fails to his gain if it happens.

Prop. 3

The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2nd on supposition the 1st happens.

Suppose that, if both events happen, I am to receive N, that the probability both will happen is P/N, that the 1st will is a/N (and consequently that the 1st will not is $(N-a)/N$) and that the 2nd will happen upon supposition the 1st does is b/N. Then (by definition 5) P will be the value of my expectation, which will become b if the 1st happens. Consequently if the 1st happens, my gain by it is $b-P$, and if it fails my loss is P. Wherefore, by the foregoing proposition, a/N is to $(N-a)/N$, i.e. a is to $N-a$ as P is to $b-P$. Wherefore (*componendo inverse*) a is to N as P is to b. But the ratio of P to N is compounded of the ratio of P to b, and that of b to N. Wherefore the same ratio of P to N is compounded of the ratio of a to N and that of b to N, i.e. the probability that the two subsequent events will both happen is compounded of the probability of the 1st and the probability of the 2nd on supposition the 1st happens.

COROLLARY. Hence if of two subsequent events the probability of the 1st be a/N, and the probability of both together be P/N, then the probability of the 2nd on supposition the 1st happens is P/a.

Prop. 4

If there be two subsequent events to be determined every day, and each day the probability of the 2nd is b/N and the probability of both P/N, and I am to receive N if both the events happen the first day on which the 2nd does; I say, according to these conditions, the probability of my obtaining N is P/b. For if not, let the probability of my obtaining N be x/N and let y be to x as $N-b$ to N. Then since x/N is the probability of my obtaining N (by definition 1) x is the value of my expectation. And again, because according to the foregoing conditions the first day I have an expectation of obtaining N depending on the happening of both the events together, the probability of which is P/N, the value of this expectation is P. Likewise, if this coincident should not happen I have an expectation of being reinstated in my former circumstances, i.e. of receiving that which in value is x depending on the failure of the 2nd event the probability of which (by cor. prop. 1) is $(N-b)/N$ or y/x, because y is to x as $N-b$ to N. Wherefore since x is the thing expected and y/x the probability of obtaining it, the value of this expectation is y. But these two last expectations together are evidently the same with my original expectation, the value of which is x, and therefore $P+y = x$. But y is to x as $N-b$ is to N. Wherefore x is to P as N is to b, and x/N (the probability of my obtaining N) is P/b.

COR. Suppose after the expectation given me in the foregoing proposition, and before it is at all known whether the 1st event has happened or not, I should find that the 2nd event has happened; from hence I can only infer that the event is determined on which my expectation depended, and have no reason to esteem the value of my expectation either greater or less than it was before. For if I have reason to think it less, it would be reasonable for me to give something to be reinstated in my former circumstances, and this over and over again as often as I should be informed that the 2nd event had happened, which is evidently absurd. And the like absurdity plainly follows if you say I ought to set a greater value on my expectation than before, for then it would be reasonable for me to refuse something if offered me upon condition I would relinquish it, and be reinstated in my former circumstances; and this likewise over and over again as often as (nothing being known concerning the 1st event) it should appear that the 2nd had happened. Notwithstanding therefore this discovery that the 2nd event has happened, my expectation ought to be esteemed the same in value as before, i.e. x, and consequently the probability of my obtaining N is (by definition 5) still x/N or P/b.* But after this discovery the probability of my obtaining N is the probability that the 1st of two subsequent events has happened upon the supposition that the 2nd has, whose probabilities were as before specified. But the probability that an event has happened is the same as the probability I have to guess right if I guess it has happened. Wherefore the following proposition is evident.

* What is here said may perhaps be a little illustrated by considering that all that can be lost by the happening of the 2nd event is the chance I should have had of being reinstated in my former circumstances, if the event on which my expectation depended had been determined in the manner expressed in the proposition. But this chance is always as much *against* me as it is *for* me. If the 1st event happens, it is *against* me, and equal to the chance for the 2nd event's failing. If the 1st event does not happen, it is *for* me, and equal also to the chance for the 2nd event's failing. The loss of it, therefore, can be no disadvantage.

Prop. 5

If there be two subsequent events, the probability of the 2nd b/N and the probability of both together P/N, and it being first discovered that the 2nd event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b.*

Prop. 6

The probability that several independent events shall all happen is a ratio compounded of the probabilities of each.

For from the nature of independent events, the probability that any one happens is not altered by the happening or failing of any of the rest, and consequently the probability that the 2nd event happens on supposition the 1st does is the same with its original probability; but the probability that any two events happen is a ratio compounded of the probability of the 1st event, and the probability of the 2nd on supposition the 1st happens by prop. 3. Wherefore the probability that any two independent events both happen is a ratio compounded of the probability of the 1st and the probability of the 2nd. And in like manner considering the 1st and 2nd events together as one event; the probability that three independent events all happen is a ratio compounded of the probability that the two 1st both happen and the probability of the 3rd. And thus you may proceed if there be ever so many such events; from whence the proposition is manifest.

Cor. 1. If there be several independent events, the probability that the 1st happens the 2nd fails, the 3rd fails and the 4th happens, etc. is a ratio compounded of the probability of the 1st, and the probability of the failure of the 2nd, and the probability of the failure of the 3rd, and the probability of the 4th, etc. For the failure of an event may always be considered as the happening of its contrary.

Cor. 2. If there be several independent events, and the probability of each one be a, and that of its failing be b, the probability that the 1st happens and the 2nd fails, and the 3rd fails and the 4th happens, etc. will be $abba$, etc. For, according to the algebraic way of notation, if a denote any ratio and b another, $abba$ denotes the ratio compounded of the ratios a, b, b, a. This corollary therefore is only a particular case of the foregoing.

Definition. If in consequence of certain data there arises a probability that a certain event should happen, its happening or failing, in consequence of these data, I call it's happening or failing in the 1st trial. And if the same data be again repeated, the happening or failing of the event in consequence of them I call its happening or failing in the 2nd trial; and so on as often as the same data are repeated. And hence it is manifest that the happening or failing of the same event in so many diffe[rent] trials, is in reality the happening or failing of so many distinct independent events exactly similar to each other.

* What is proved by Mr Bayes in this and the preceding proposition is the same with the answer to the following question. What is the probability that a certain event, when it happens, will be accompanied with another to be determined at the same time? In this case, as one of the events is given, nothing can be due for the expectation of it; and, consequently, the value of an expectation depending on the happening of both events must be the same with the value of an expectation depending on the happening of one of them. In other words; the probability that, when one of two events happens, the other will, is the same with the probability of this other. Call x then the probability of this other, and if b/N be the probability of the given event, and p/N the probability of both, because $p/N = (b/N) \times x$, $x = p/b$ = the probability mentioned in these propositions.

On, then, to Bayes' **Essay toward solving a problem in the doctrine of chances** My comments are interwoven with the actual text. Bayes starts with a series of definitions, of which only the rather tautological sounding Definition 5 seems to need some commenting on. The others define disjunction (of two and more events) and independence.

Definition 5: the probability of an event is defined as the ratio between two "values": (1) that at which "an expectation depending on the happening of an event ought to be computed"; and (2) that "of the thing expected upon its happening."

So, let us assume an event A, upon whose happening I am to receive some value N, this value is the "thing" I expect upon the event A's happening. (But A may not happen, in which case I get no-thing.)

But what do we understand by the numerator, that is, the value at which the expectation depending on A's happening *ought to be computed*? Say, I hazard a, the maximum amount I am willing to pay for receiving either N (if A occurs) or nothing (if A doesn't occur). This means treating a as something like my just desserts – which is why I should be rationally willing to hazard that amount. So, Definition 5 says that P(A) = $\frac{a}{N}$ or a = P(A)*N. Under this interpretation, since a is what I "ought" to get, if A happens, then I gain *(N-a)*, and if it doesn't I lose a (see Proposition 2 below). Given these conditions, the game is equitable in the sense that my average gain or loss is 0. That is,

$$P(A) * (gain) + P(\neg A) * (loss) = \frac{a}{N} * (N - a) + \frac{N - a}{N} * (-a) = 0 \tag{3}$$

so that while I may win or lose certain amounts, the game itself is not biased against me: winning once and losing once returns me to my initial situation. If one were taking a frequentist view of the game, then the definition implies that in the long run, if I bet a each time, I will neither gain nor lose anything.

Hence, a represents the value at which my expectation "ought to be computed." But the idea of being willing to pay a in order to participate in the game is not purely subjective; rather, the phrase "ought to be computed" implies an objective understanding, perhaps something akin to what John Earman has called "rational or justified degree of belief," and this can be computed.

Propostion 1: essentially establishes the so-called "additive axiom." That is, given 3 disjunct events A, B, and C, with respective probabilities of occurrence $\frac{a}{N}$, $\frac{b}{N}$, $\frac{c}{N}$, then the probability of the union of A, B, and C (that is, of either A or B or C occurring) is given by

$\frac{a}{N} + \frac{b}{N} + \frac{b}{N}$.

To establish this, Bayes uses Definition 5 in the form a = P(A)*N. So, for the events A, B, and C, the values at which the expectation of the events' happening "ought to be computed" are a, b, and c respectively. Thus the total expectation from all three is a + b + c. But if any of these three happens, I get N, so, according Definition 5 (in its form: P(A) = $\frac{a}{N}$),

$$P(A\ or\ B\ or\ C) = \frac{a+b+c}{N} = \frac{a}{N} + \frac{b}{N} + \frac{c}{N} \tag{4}$$

The two corollaries that follow simply establish:

(1) that if A, B, and C partition the universe Ω, then the total value at which the sum total of expectations ought to be computed (a + b + c) = N. Since one of these events must happen, and I get N upon its happening, I am certain of getting N. In this case, value of thing expected upon the happening (of either A or B or C) must equal the value at which the event (of either A or B or C) ought to be computed.

(2) that, for any event, the probability $P(\neg A) = 1 - P(A)$, so that if

$$P(A) = \frac{a}{N}$$

then

$$P(\neg A) = \frac{N-a}{N}$$

.

Proposition 2 establishes a relationship between gains and losses that Bayes will use repeatedly in the subsequent propositions.

Let P(A) = $\frac{p}{N}$. This means, the value at which the expectation ought to be computed is p – these are my just desserts. If A happens, I "gain" (N-p), whereas if it fails, I lose p. So, the ratio of loss if it fails to my gain if it succeeds is:

$$\frac{p}{N-p}$$

But this is equal to

$$\frac{\frac{p}{N}}{\frac{N-p}{N}}$$

The numerator is simply P(A), while the denominator is P(¬A), so the ratio of expected loss to expected gain is equal to the ratio between the probability of event's success and the probability of event's failure, that is,

$$\frac{p}{N-p} = \frac{P(A)}{P(\neg A)} \tag{5}$$

Proposition 4: with this (and the subsequent proposition 5), we come near the heart of Bayes' contribution to the history of probability – the problem of inference or inverse probability, which will be fundamental to his thought-experiment to follow.

Proposition 4 offers us the inverse of Proposition 3. That is, given two successive events, A and B, the probability of my inferring that the first has occurred *given that the second has occurred* is the ratio of the unconditional probability of both A and B occurring to the unconditional probability of B occurring. In modern notation:

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

Now, I simply have been unable to re-construct (as I did for Proposition 3) the logic of Bayes' demonstration here. He begins with the known quantities: $P(B) = \frac{b}{N}$ and $P(A \cap B) = \frac{p}{N}$, where N is what I receive if both A and B occur (that is, "the value of the thing expected upon its happening," p being the "rational" expectation or the rate at which expectation ought to be computed). He sets out then to derive my probability of obtaining N, under the condition that B has occurred, that is, to show that

$$P(A|B) = \frac{\frac{p}{N}}{\frac{b}{N}} = \frac{p}{b}$$

But I find the description of the procedure for doing so utterly obscure, and since all we need is the result, I won't pursue this further.

Proposition 5 does not offer a new result but is crucial as an interpretation of what Proposition 4 has established: that is, it lays out the idea of inductive inference. If proposition 4 had laid out the idea that we can compute the probability of a prior event having occurred based upon the occurring of a subsequent event, Proposition 5 says that this computation yields the probability of my being right when, upon seeing that B has occurred, I guess that the prior event A must have occurred. As he says at the end of Proposition 4, "the probability that an event has happened is the same as the the probability [that I] guess right if I guess it has happened," and consequently what we have computed in Proposition 4 simply measures the likelihood that my inference (based on the happening of a subsequent event) regarding the prior event's happening is correct.

Proposition 6 focuses on independence, that is, on a series of events (for instance, a sequence of coin tosses) where the occurrence of any one event has no bearing on the occurrence of another, or, as Bayes puts it, "the probability that any one happens is not altered by the happening or failing of any of the rest." The proposition states that for a series of independent events the probability that they all occur is simply the multiplication of their individual probabilities, that is,

$$P(A \cap B \cap C \ldots) = P(A) * P(B) * P(C) \ldots$$

Bayes takes as given in the course of arguing for this that the claim holds for any two events, that is, $P(A \cap B) = P(A) * P(B)$. Bayes provides, too, the modern definition of independence, though taking this as evident: "the probability that the second event happens on the supposition the first does is the same with its original probability," or

$$P(B|A) = P(B)$$

These two claims are actually intertwined, for

$$P(A \cap B) = P(A) * P(B) \Rightarrow P(B) = \frac{P(A \cap B)}{P(A)} = P(B|A)\ (by\ Proposition\ 3)$$

Proposition 6 extends the two event case to several independent events. Given three events A, B, and C, we want $P(A \cap B \cap C)$. We can treat this as the conjunction of two events: (1) $A \cap B$ or the event of both A and B occurring, and (2) the event C. Now, the probability of $A \cap B$ is simply P(A)*P(B), hence the probability of all three events occurring is simply P(A)*P(B)*P(C), and so on. The corollary extends this to a mixture of successes and failures by treating the failure of an event to happen as "the happening of its contrary."

Proposition 7: links the repetition of an event (with a known probability of failure and success in a single trial) to the binomial theorem. It says that if $P(A) = a$ and $P(\neg A) = b$, then the probability of A happening exactly p times and failing exactly q times in (p+q) trials is given by the corresponding term in the binomial expansion of $(a + b)^{p+q}$. In other words,

$$P(p\ successes\ and\ q\ failures) = \binom{p+q}{p} a^p b^q$$

The important assumption here is that repetitions of the same event are seen as being independent of one another. With all these ideas in place, we are finally ready to turn to:

Prop. 7

If the probability of an event be a, and that of its failure be b in each single trial, the probability of its happening p times, and failing q times in $p+q$ trials is Ea^pb^q if E be the coefficient of the term in which occurs a^pb^q when the binomial $(a+b)^{p+q}$ is expanded.

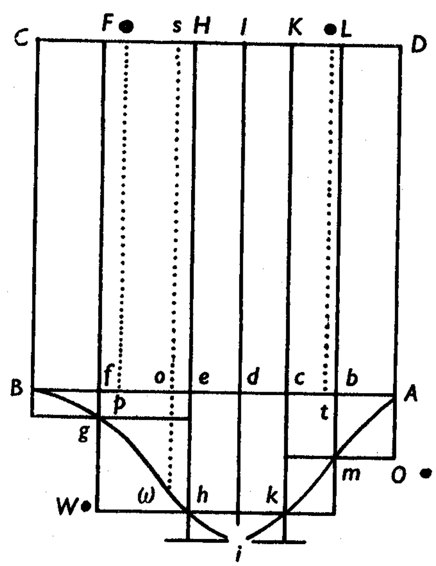
For the happening or failing of an event in different trials are so many independent events. Wherefore (by cor. 2 prop. 6) the probability that the event happens the 1st trial, fails the 2nd and 3rd, and happens the 4th, fails the 5th, etc. (thus happening and failing till the number of times it happens be p and the number it fails be q) is $abbab$ etc. till the number of a's be p and the number of b's be q, that is; 'tis a^pb^q. In like manner if you consider the event as happening p times and failing q times in any other particular order, the probability for it is a^pb^q; but the number of different orders according to which an event may happen or fail, so as in all to happen p times and fail q, in $p+q$ trials is equal to the number of permutations that $aaaa\ bbb$ admit of when the number of a's is p, and the number of b's is q. And this number is equal to E, the coefficient of the term in which occurs a^pb^q when $(a+b)^{p+q}$ is expanded. The event therefore may happen p times and fail q in $p+q$ trials E different ways and no more, and its happening and failing these several different ways are so many inconsistent events, the probability for each of which is a^pb^q, and therefore by prop. 1 the probability that some way or other it happens p times and fails q times in $p+q$ trials is Ea^pb^q.

Section II

Postulate. 1. I suppose the square table or plane $ABCD$ to be so made and levelled, that if either of the balls o or W be thrown upon it, there shall be the same probability that it rests upon any one equal part of the plane as another, and that it must necessarily rest somewhere upon it.

2. I suppose that the ball W shall be first thrown, and through the point where it rests a line os shall be drawn parallel to AD, and meeting CD and AB in s and o; and that afterwards the ball O shall be thrown $p+q$ or n times, and that its resting between AD and os after a single throw be called the happening of the event M in a single trial. These things supposed:

Lem. 1. The probability that the point o will fall between any two points in the line AB is the ratio of the distance between the two points to the whole line AB.



Let any two points be named, as f and b in the line AB, and through them parallel to AD draw fF, bL meeting CD in F and L. Then if the rectangles Cf, Fb, LA are commensurable to each other, they may each be divided into the same equal parts, which being done, and the ball W thrown, the probability it will rest somewhere upon any number of these equal parts will be the sum of the probabilities it has to rest upon each one of them, because its resting upon any different parts of the plane AC are so many inconsistent events; and this sum, because the probability it should rest upon any one equal part as another is the same, is the probability it should rest upon any one equal part multiplied by the number of parts. Consequently, the probability there is that the ball W should rest somewhere upon Fb is the probability it has to rest upon one equal part multiplied by the number of equal parts in Fb; and the probability it rests somewhere upon Cf or LA, i.e. that it does not rest upon Fb (because it must rest somewhere upon AC) is the probability it rests upon one equal part multiplied by the number of equal parts in Cf, LA taken together. Wherefore, the probability it rests upon Fb is to the probability it does not as the number of equal parts in Fb is to the number of equal parts in Cf, LA together, or as Fb to Cf, LA together, or as fb to Bf, Ab together. Wherefore the probability it rests upon Fb is to the probability it does not as fb to Bf, Ab together. And (*componendo inverse*) the probability it rests upon Fb is to the probability it rests upon Fb added to the probability it does not, as fb to AB, or as the ratio of fb to AB to the ratio of AB to AB. But the probability of any event added to the probability of its failure is the ratio of equality; wherefore, the probability it rests upon Fb is to the ratio of equality as the ratio of fb to AB to the ratio of AB to AB, or the ratio of equality; and therefore the probability it rests upon Fb is the ratio of fb to AB. But *ex hypothesi* according as the ball W falls upon Fb or not the point o will lie between f and b or not, and therefore the probability the point o will lie between f and b is the ratio of fb to AB.

Again; if the rectangles Cf, Fb, LA are not commensurable, yet the last mentioned probability can be neither greater nor less than the ratio of fb to AB; for, if it be less, let it be the ratio of fc to AB, and upon the line fb take the points p and t, so that pt shall be greater than fc, and the three lines Bp, pt, tA commensurable (which it is evident may be always done by dividing AB into equal parts less than half cb, and taking p and t the nearest points of division to f and c that lie upon fb). Then because Bp, pt, tA are commensurable, so are the rectangles Cp, Dt, and that upon pt compleating the square AB. Wherefore, by what has been said, the probability that the point o will lie between p and t is the ratio of pt to AB. But if it lies between p and t it must lie between f and b. Wherefore, the probability it should lie between f and b cannot be less than the ratio of pt to AB, and therefore must be greater than the ratio of fc to AB (since pt is greater than fc). And after the same manner you may prove that the forementioned probability cannot be greater than the ratio of fb to AB, it must therefore be the same.

LEM. 2. The ball W having been thrown, and the line os drawn, the probability of the event M in a single trial is the ratio of Ao to AB.

For, in the same manner as in the foregoing lemma, the probability that the ball o being thrown shall rest somewhere upon Do or between AD and so is the ratio of Ao to AB. But the resting of the ball o between AD and so after a single throw is the happening of the event M in a single trial. Wherefore the lemma is manifest.

Prop. 8

If upon BA you erect the figure $BghikmA$ whose property is this, that (the base BA being divided into any two parts, as Ab, and Bb and at the point of division b a perpendicular being erected and terminated by the figure in m; and y, x, r representing respectively the ratio of bm, Ab, and Bb to AB, and E being the coefficient of the term in which occurs $a^p b^q$ when the binomial $(a+b)^{p+q}$ is expanded) $y = Ex^p r^q$. I say that before the ball W is thrown, the probability the point o should fall between f and b, any two points named in the line AB, and withall that the event M should happen p times and fail q in $p+q$ trials, is the ratio of $fghikmb$, the part of the figure $BghikmA$ intercepted between the perpendiculars fg, bm raised upon the line AB, to CA the square upon AB.

within the figure, as eg, dh, dk, cm, you may prove that the last mentioned probability is *greater* than the ratio of any figure less than $fghikmb$ to CA.

Wherefore, that probability must be the ratio of $fghikmb$ to CA.

Cor. Before the ball W is thrown the probability that the point o will lie somewhere between A and B, or somewhere upon the line AB, and withal that the event M will happen p times, and fail q in $p+q$ trials is the ratio of the whole figure AiB to CA. But it is certain that the point o will lie somewhere upon AB. Wherefore, before the ball W is thrown the probability the event M will happen p times and fail q in $p+q$ trials is the ratio of AiB to CA.

Prop. 9

If before anything is discovered concerning the place of the point o, it should appear that the event M had happened p times and failed q in $p+q$ trials, and from hence I guess that the point o lies between any two points in the line AB, as f and b, and consequently that the probability of the event M in a single trial was somewhere between the ratio of Ab to AB and that of Af to AB: the probability I am in the right is the ratio of that part of the figure AiB described as before which is intercepted between perpendiculars erected upon AB at the points f and b, to the whole figure AiB.

For, there being these two subsequent events, the first that the point o will lie between f and b; the second that the event M should happen p times and fail q in $p+q$ trials; and (by cor. prop. 8) the original probability of the second is the ratio of AiB to CA, and (by prop. 8) the probability of both is the ratio of $fghimb$ to CA; wherefore (by prop. 5) it being first discovered that the second has happened, and from hence I guess that the first has happened also, the probability I am in the right is the ratio of $fghimb$ to AiB, the point which was to be proved.

Cor. The same things supposed, if I guess that the probability of the event M lies somewhere between 0 and the ratio of Ab to AB, my chance to be in the right is the ratio of Abm to AiB.

Scholium

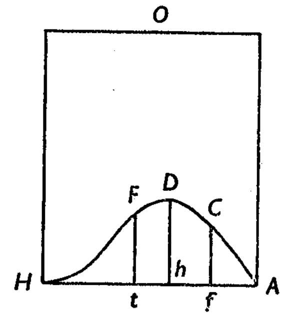
From the preceding proposition it is plain, that in the case of such an event as I there call M, from the number of times it happens and fails in a certain number of trials, without knowing anything more concerning it, one may give a guess whereabouts it's probability is, and, by the usual methods computing the magnitudes of the areas there mentioned, see the chance that the guess is right. And that the same rule is the proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. For, on this account, I may justly reason concerning it as if its probability had been at first unfixed, and then determined in such a manner as to give me no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. But this is exactly the case of the event M. For before the ball W is thrown, which determines it's probability in a single trial (by cor. prop. 8), the probability it has to happen p times and fail q in $p+q$ or n trials is the ratio of AiB to CA, which ratio is the same when $p+q$ or n is given, whatever number p is; as will appear by computing the magnitude of AiB by the method of fluxions.* And consequently before the place of the point o is discovered or the number of times the event M has happened in n trials, I can have no reason to think it should rather happen one possible number of times than another.

In what follows therefore I shall take for granted that the rule given concerning the event M in prop. 9 is also the rule to be used in relation to any event concerning the probability of which nothing at all is known antecedently to any trials made or observed concerning it. And such an event I shall call an unknown event.

Cor. Hence, by supposing the ordinates in the figure AiB to be contracted in the ratio of E to one, which makes no alteration in the proportion of the parts of the figure intercepted between them, and applying what is said of the event M to an unknown event, we have the following proposition, which gives the rules for finding the probability of an event from the number of times it actually happens and fails.

Prop. 10

If a figure be described upon any base AH (Vid. Fig.) having for it's equation $y = x^p r^q$; where y, x, r are respectively the ratios of an ordinate of the figure insisting on the base at right angles, of the segment of the base intercepted between the ordinate and A the beginning of the base, and of the other segment of the base lying between the ordinate and the point H, to the base as their common consequent. I say then that if an unknown event has happened p times and failed q in $p+q$ trials, and in the base AH taking any two points as f and t you erect the ordinates fC, tF at right angles with it, the chance that the probability of the event lies somewhere between the ratio of Af to AH and that of At to AH, is the ratio of $tFCf$, that part of the before-described figure which is intercepted between the two ordinates, to $ACFH$ the whole figure insisting on the base AH.



This is evident from prop. 9 and the remarks made in the foregoing scholium and corollary.

Now, in order to reduce the foregoing rule to practice, we must find the value of the area of the figure described and the several parts of it separated, by ordinates perpendicular to its base. For which purpose, suppose $AH = 1$ and HO the square upon AH likewise $= 1$, and Cf will be $= y$, and $Af = x$, and $Hf = r$, because y, x and r denote the ratios of Cf, Af, and Hf respectively to AH. And by the equation of the curve $y = x^p r^q$ and (because $Af + fH = AH$) $r + x = 1$. Wherefore

$$y = x^p(1-x)^q$$
$$= x^p - qx^{p+1} + \frac{q(q-1)x^{p+2}}{2} - \frac{q(q-1)(q-2)x^{p+3}}{2.3} + \text{etc.}$$

Now the abscisse being x and the ordinate x^p the correspondent area is $x^{p+1}/(p+1)$ (by prop. 10, cas. 1, Quadrat. Newt.)† and the ordinate being qx^{p+1} the area is $qx^{p+2}/(p+2)$; and

* It will be proved presently in art. 4 by computing in the method here mentioned that AiB contracted in the ratio of E to 1 is to CA as 1 to $(n+1)E$: from whence it plainly follows that, antecedently to this contraction, AiB must be to CA in the ratio of 1 to $n+1$, which is a constant ratio when n is given, whatever p is.

† 'Tis very evident here, without having recourse to Sir Isaac Newton, that the fluxion of the area ACf being

$$y\dot{x} = x^p\dot{x} - qx^{p+1}\dot{x} + \frac{q(q-1)}{2}x^{p+2}\dot{x} - \text{etc.}$$

the fluent or area itself is
$$\frac{x^{p+1}}{p+1} - \frac{qx^{p+2}}{p+2} + \frac{q(q-1)x^{p+3}}{2(p+3)} - \text{etc.}$$

Bayes' Gedanken-Experiment

The Problem: Consider an event M that has a certain probability of occurring in a single trial (say, θ). I then run (p + q) trials and the event M occurs p times. What Bayes seeks is the following: given that the event M occurred p times, with what certainty can we reconstruct the probability of a single occurrence? Or, as Price puts it, what is the probability that its probability of occurrence in a single trial (θ) lies between "any two degrees of probability that can be named," under the condition that it has occurred p times in the past. In short, for any two probability values a and b that we choose, we require

$$P(a < \theta < b|M = p)$$

The Model: To figure this out, Bayes imagines a square table upon which a ball W is rolled (see his diagram). Wherever it comes to rest, a vertical line is drawn through that point, parallel to the sides of the square. Thereafter, another ball is thrown repeatedly in the same manner (p + q) = n times, and a vertical line is drawn as above. Each time the ball will stop either to the right or to the left of the line upon which W came to rest. For each single trial, if it falls to the right, this is the happening of the event M (if it stops on the left, it is the event of M not-happening).

Bayes has thus defined two events that are to be related– (a) the initial event of rolling W, and then (b) a second event dependent upon this first, that of the p + q trials of rolling another ball and noting on how many occasions – p – it stopped to the right of W.

So, the problem boils down to the following: if we know that M happened p times (that is, we know that the second event occurred), what is the probability that the probability θ of the first event (rolling of the ball W) lies between any two probabilities we name (call them θ_1 and θ_2). Or, how good is our guess at a particular θ for the event of W?

Using Propositions 4 and 5 above, we can say that

$$P(\theta_1 < \theta < \theta_2|M = p) = \frac{P[(\theta_1 < \theta < \theta_2) \cap (M = p)]}{P(M = p)} \tag{6}$$

Lemma 1 allows Bayes to specify probabilities of the ball's position spatially in relation to length of the square's side. He shows that probabilities corresponding to where a ball stops are simply given by ratios of lengths of line segments to the length of the table's side.

While complicated in its exegesis, the idea here is quite straightforward. Consider dividing the table into a number of smaller rectangles, each of whose vertical sides is equal in length to the side of the table. Intuitively, the probability that a ball will come to rest within a particular rectangle is simply the ratio of the area of that rectangle to the area of table as a whole. But since the lengths of the vertical sides of all the rectangles are the same and equal to that of the table's side, this probability will reduce to the ratio of the width of a particular rectangle to the length of the table's side.

Thus, to say that θ lies between two named probabilities is to say that W comes to rest upon a line that falls upon a specified segment of the table's base. The ratio of the lengths of the two endpoints of the segment to the length of the table's side provides the two named probabilities.

Lemma 2 establishes the probability relationships between the two events (W and M = p). As he shows, the initial event – the throwing of the ball W – defines the probability in a single trial of each subsequent throw falling to the left or right of W. That is, the first event specifies a θ for all subsequent trials, since for each subsequent trial the probability of M happening in a single trial (that is, of a ball stopping to the right of W) is simply the ratio of the length of the table's side to the right of where W stopped to the full length of the table's side. In other words, for a single trial:

$$P(M) = \frac{portion\ of\ side\ to\ W's\ right}{length\ of\ side}\ likewise,\ P(\neg M) = \frac{portion\ of\ side\ to\ W's\ left}{length\ of\ side}$$

Consequences of these Lemmas: Probability distribution of second event is Binomial.

As Bayes points out, to specify the probability in a single trial of M happening means that the event of (p + q) trials can be treated via the binomial expansion theorem. Think of each trial of the event M as a coin toss, with some probability θ of heads (falling to the right of W) and $(1-\theta)$ of tails (falling to the left). So, toss the coin (p + q) times, getting p heads – this is the given condition. Now, we might, for instance, get the following sequence : HHHHHH . . . p times, followed by TTTTT . . . q times. Since the tosses are independent and each toss has a probability θ of heads, the probability of this particular sequence is given by $\theta^p(1-\theta)^q$. But there are $\binom{p+q}{p}$ ways of getting p heads, so the total probability of getting p heads (that is, of the ball falling to the right p times) is $\binom{p+q}{p}\theta^p(1-\theta)^q$. We recognise in this the familiar binomial expansion (see Proposition 7).

To sum up, given W's position, if in (p + q) trials the ball ends up to the right of W p times, the likelihood of this happening is given by corresponding term in the binomial expansion of $(\theta + (1-\theta))^{p+q}$:

$$P(M = p|\theta) = \binom{p+q}{p}\theta^p(1-\theta)^q \tag{7}$$

Using these two lemmas and his earlier propositions, Bayes can now approach the problem of inverse probability, and work out the conditional probability of W's position, given that we know M = p. To do this he needs to work out two unconditional probabilities: the unconditional probability of M = p *over all possible positions W can take* and the unconditional joint probability of M = p and of W falling between two specified values. I will first lay out Bayes' analysis in modern terms before repeating the analysis in terms of his spatial model.

How does he calculate these unconditional probabilities?

Now, the equation above is true given a *fixed* θ. But θ can range from 0 to 1, corresponding to all the different places that the ball W can come to rest in that initial throw. So, to find the *total* probability of M = p for *all* possible positions of W, we have to integrate equation 5 over all values of θ. Consequently,

$$P(M = p) = \int_0^1 \binom{p+q}{p}\theta^p(1-\theta)^q \, d\theta \tag{8}$$

The same logic yields the other element we also need to know: the joint probability that M happens p times in (p + q) trials *and* the initial roll of W falls in some specified segment of the table's base. That is, we need to know the probability that θ ranges from θ_1 to θ_2, say, *and* M = p, where θ_1 and θ_2 are simply the respective ratios of the lengths of the end-points of the segment to the table's side (from Lemma 1). All we need to do to obtain this is to integrate the right-hand term in equation 7 from θ_1 to θ_2 (rather than from 0 to 1), yielding:

$$P[(\theta_1 < \theta < \theta_2) \cap (M = p)] = \int_{\theta_1}^{\theta_2} \binom{p+q}{p}\theta^p(1-\theta)^q \, d\theta \tag{9}$$

Combining equations 8 and 9 with the result of Proposition 4 above, we get the desired answer:

$$P(\theta_1 < \theta < \theta_2 | M = p) = \frac{\displaystyle\int_{\theta_1}^{\theta_2} \binom{p+q}{p} \theta^p (1-\theta)^q \, d\theta}{\displaystyle\int_0^1 \binom{p+q}{p} \theta^p (1-\theta)^q \, d\theta} \tag{10}$$

In other words, given that (p + q) trials have produced p successes (so M = p), our belief that the probability of success in a single trial lies between two specified probabilities a and b is obtained by the ratio of two "areas": (1) the area representing the probability that p successes occur when the probability of success in a single trial is between θ_1 and θ_2; and (2) the area representing the probability that p successes occur when the probability of success in single trial ranges over all possible values from 0 to 1.

So, what do Bayes' actual propositions look like?

Proposition 8 and its corollary:

Let me choose a point – which Bayes calls b – on the base of the square such that the ball W came to a rest somewhere vertically in line with that point. As we have seen, the probability that any subsequent throw will fall to the right of this W is $\theta = \frac{Ab}{AB}$, where AB is the length of the base that extends from A to B.

Consequently, the probability that p of the p + q subsequent throws is to the right of M (i.e, $P(M = p|\theta)$ is $\binom{p+q}{p}\theta^p(1-\theta)^q$

Let me now draw a figure on the base corresponding to each possible specification of θ, that is, corresponding to each possible point b. At each possible point b, the figure is defined to have a height (bm) which we calculate by multiplying the length of the base (AB) with the probability (given M = p) corresponding to that b. That is,

$$bm = AB * \binom{p+q}{p}\theta^p(1-\theta)^q \tag{11}$$

where $\theta = \frac{Ab}{AB}$. If theta ranges from 0 to 1 – that is, W's position projected onto AB (i.e. b) ranges over the range of all possible values it can take from 0 to AB – then each value of θ specifies the height of this figure (bm) at each point from A to B.

What would the shape of this figure be like? Clearly, if $\theta = 0$, then the equation above yields bm = 0, intuitively true since if the line where W falls is at the far right, there can be no subsequent roll that will fall to the right of that line. Likewise, if $\theta = 1$, then bm = 0, since this would mean that W falls on the far left side of the table, and all subsequent trials will result in the ball falling to its right, which means that there is no chance of our achieving exactly p successes. And the peak? Well, this will depend upon the ratio of successes of M to the total number of trials. One can show that the figure will peak at a $\theta = \frac{p}{p+q}$. So, for instance, if we have p = q, so that the number of successes is equal to exactly half the number of trials, then the the figure will be symmetric, peaking at the midpoint of AB and tapering down to 0 at the endpoints A and B.

Now consider what the area under this curve represents.

The area of the square table (AB^2) represents the universe of outcomes Ω. In other words, the total area is proportional to all the possible successes of M (that is, M ranges from 0 to p + q) over all the possible positions of W (θ ranges from 0 to 1). By proportional

I mean the following: the probability of a ball falling in any sector of the table is given by the ratio of the area of that sector to the area of the whole table. A sector represents the total number of outcomes for all possible values of M (from 0 to p + q) over a particular range of θ (the range depends upon width of that sector with respect to the length of the side of the table.

Now, if we are given M = p, then we focus on a *subset* of these outcomes: namely, just those of the p + q trials in which we had p successes. That is, for each possible position of W (θ ranges from 0 to 1), we restrict ourself only to those trials where the ball ended up to the right of W on p occasions. This is the *unconditional* probability of M = p in p + q trials. It is precisely this that is proportional to the area under the curve. Now the area under the figure is obtained by integrating bm over all possible values of θ – think of integration as simply adding up the different heights bm that correspond to each value of θ. In short, we have the equation:

$$P[M = p] = \frac{AB * \int_0^1 \binom{p+q}{p} \theta^p (1-\theta)^q \, d\theta}{AB^2} \tag{12}$$

which represents the probability of a subset of outcomes over all possible positions of W, for a given M = p. The denominator is simply the area of the universe Ω of all possible outcomes (the ratio of the two areas gives the probability). To sum up: the (unconditional) probability of this subset occurring is therefore simply obtained by dividing the result of this integral by the area representing Ω, that is, AB^2. This result provides Bayes' corollary to Proposition 8.

The actual proposition 8 requires us to consider the probability of an even smaller subset of outcomes. If we are given M = p *and* $(b < W < f)$, we further restrict ourselves to *only* those p-success trials for *only* those θ values *corresponding to W lying between b and f.* This is the (unconditional) joint probability of these two events relative to Ω. Now, if integrating the curve over all values of θ allowed us to specify the total probability of M = p regardless of where W fell (that is, for all positions of W), then integrating the same curve over the θ values corresponding to a restricted set of W positions (between b and f) and then dividing by the area of Ω will yield the probability we want. And this is the result of Proposition 8.

$$P[(M = p) \cap b < W < f] = \frac{AB * \int_{\theta_1}^{\theta_2} \binom{p+q}{p} \theta^p (1-\theta)^q \, d\theta}{AB^2} \tag{13}$$

where θ_1 and θ_2 represent the θ values corresponding to the positions b and f respectively, between which W is allowed to range.

Proposition 9: The final result we want

Now, we can derive the conditional probability we want quite easily using Propositions 4 and 5. Recall from them that the conditional probability of A given B was shown to be the ratio of the unconditional probability of A and B divided by the unconditional probability of B. That is,

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

Here, we want the conditional probability that W fell between any two chosen points on the line AB, given that M = p. This is simply the unconditional probability of both these events divided by the unconditional probability of M = p. Since $b < W < f$ is equivalent to $\theta_1 < \theta < \theta_2$, we can write this as

$$P[(\theta_1 < \theta < \theta_2)|M = p] = \frac{P[M = p \cap (\theta_1 < \theta < \theta_2)]}{P[M = p]}$$

But the numerator and denominator of this are readily available from the equations 12 and 13 above, yielding the result:

$$P[(\theta_1 < \theta < \theta_2)|M = p] == \frac{\displaystyle\int_{\theta_1}^{\theta_2} \binom{p+q}{p} \theta^p (1-\theta)^q \, d\theta}{\displaystyle\int_0^1 \binom{p+q}{p} \theta^p (1-\theta)^q \, d\theta} \tag{14}$$

As the *Scholium* says, this proposition establishes "that in the case of such an event as I there call M, from the number of times it happens and fails in a certain number of trials, without knowing anything more concerning it, one may give a guess where its probability is, and ...see the chance that the guess is right." In other words, knowing how many times something has happened or failed allows to compute how likely we are to be correct in our guess, if we say that the event W occurred within any certain range we guess at. The Scholium has more to this, and it is worth looking at closely.