The poverty of Venn diagrams for teaching probability: their history and replacement by Eikosograms

W.H. Cherry and R.W. Oldford
University of Waterloo

September 23, 2002

Abstract

Diagrams convey information, some intended some not. A history of the information content of ringed diagrams and their use by Euler and Venn is given. It is argued that for the purposes of teaching introductory probability, Venn diagrams are either inappropriate or inferior to other diagrams. A diagram we call an eikosogram is shown to be coincident with what is meant by probability and so visually introduces all the rules of probability including Bayes' theorem and the product rule for independent events. Eikosograms clearly demonstrate unconditional and conditional independence – both of events and of random variables. An approach to teaching probability via the eikosogram and other more familiar diagrams is described. It is recommended that Venn diagrams no longer be used to teach probability.

Keywords: Eikosograms, Euler diagrams, Venn diagrams, outcome trees, outcome diagrams, vesica piscis, ideograms, history of probability, logic and probability, understanding conditional probability, probabilistic independence, conditional independence.

1 Introduction

It is now commonplace to use Venn diagrams to explain the rules of probability. Indeed, nearly every introductory treatment has come to rely on them. But this was not always the case. In his book *Symbolic Logic* Venn makes much use of these diagrams, yet in his book on probability, *The Logic of Chance*, they appear nowhere at all! [1]

A cursory review of some well known probability texts reveals that the first published use of these diagrams in probability may have occurred as late as 1950 with the publication of Feller's *Theory of Probability* (details are given in the Appendix). Venn diagrams don't seem to have been that much used in probability or, if used, that much appreciated. For example, Gnedenko (1966), a student of Kolmogorov, used Venn diagrams in the third edition of his text *Theory of Probability* but does not refer to them as such until the book's next edition in 1968, and then only as "so-called Venn diagrams". Even by 1969, the published use of Venn diagrams for probability was by no means common.

In more recent years, some authors of introductory probability texts have called just about any diagram which marks regions in a plane a 'Venn diagram'. Others have written that no diagram should be called a 'Venn diagram'. Dunham (1994), for example, claims that the Venn diagram was produced a century before Venn by Euler and so "If justice is to be served, we should call this an 'Euler diagram'." This view is surprisingly commonplace though not everywhere expressed as strongly as Dunham (1994 p. 262) who dismissively writes "Venn's innovation [over Euler's diagrams] . . . might just as well have been discovered by a child with a crayon." In both cases, the sense of what constitutes a Venn diagram has been lost. In the first case, the Venn diagram is not up to the job and so is stretched beyond its definition, while in the second case it is Euler's diagram that has been stretched beyond its definition to mistakenly include Venn's innovative use.

In what follows, the position is taken that diagrams convey information and like statistical graphics need to be carefully designed so as to convey the intended information and, ideally, no other. There is no all-purpose diagram; rather diagrams need to be tailored to specific purposes.

[1] It is true that Venn's probability book predates his symbolic logic book, however the diagrams only ever appeared in the latter book. This is more interesting given that Venn uses the word 'logic' in the titles of both books and also that Venn's symbolic logic used the numerical values of 1 and 0 to indicate true and false (i.e. certainty and impossibility).

In the next section we take up this point in more detail and apply it to the ringed diagrams used both by Euler and by Venn. Some history is given which demonstrates that these ringed diagrams were long in use before either Euler or Venn and would have been familiar to both men. Euler and Venn each use the diagrams in different ways as an aid to understanding logic. The diagrams have seen much use historically because they convey essentially the same information, information which is useful in many contexts. The information they convey however is not that which is most useful to teaching and understanding probability. The weaknesses of Venn diagrams for teaching probability are discussed in Section 3.

In Section 4, we explore the use of the eikosogram, a diagram which we argue is ideally suited to understanding probability. As with the ringed diagrams, the eikosogram is not a new diagram but it has not yet been put to its full use in teaching probability. Section 4 develops and uses the diagram as one would in teaching probability. From it the axioms of probability can be intuited as can conditional probability. Bayes' theorem and the subtle concepts of conditional and unconditional independence both of random variables and of events are direct consequences of, and derivable from, eikosograms.

Section 5 shows how the eikosogram complements other diagrams, notably outcome trees and outcome diagrams, to present a coordinated development of probability. The role of Venn diagrams, if it exists at all, is significantly diminished. Section 6 wraps up with some concluding remarks.

2 On Diagrams and the Meaning of Venn Diagrams

Good diagrams clarify. Very good diagrams force the ideas upon the viewer. The best diagrams compellingly embody the ideas themselves.

For example, the mathematical philosopher Ludwig Wittgenstein would have that the meaning of the symbolic expression 3×4 is had only by the "ostensive definition" shown by the diagram of Figure 1. 'What is 3×4?' can

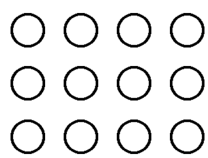

Figure 1: Defining multiplication: This figure *is* the meaning of 3×4.

exist as a question only because the diagram provides a schema for determining that $3 \times 4 = 12$. The proof of $3 \times 4 = 12$ is embodied within the definition of multiplication itself and that definition is established diagrammatically by a "perspicuous representation" (e.g. see Wittgenstein (1964) p 66 #27, p. 139, #117 or Glock, 1996, pp. 226 ff., 274 ff., 278 ff.).

Diagrams which provide ostensive definitions of fundamental mathematical concepts have a long history. In the Meno dialogue, Plato has Socrates engage in conversation with an uneducated slave boy, asking him questions about squares and triangles ultimately to arrive at the diagram in Figure 2. Although ignorant at the beginning of the dialogue,

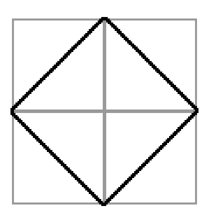

Figure 2: Each small square has area 1. The inscribed square has area of 2 and hence sides of length $\sqrt{2}$.

the slave boy comes to realize that he does indeed know how to construct a square of area 2 (the dialogue actually constructed a square of area 8, or one having sides of length $2\sqrt{2}$). Not having realized this before, nor having been told by anyone, Socrates concludes that the boy's soul must have known this from before the boy was born. With some work, the boy was able to recall this information through a series of questions. From this Socrates concludes that the soul exists and is immortal.

The simpler explanation however is that Socrates led the boy to a diagram (familiar to Socrates) which clearly shows a square of area 2. By showing the existence of the length $\sqrt{2}$, Figure 2 actually gives meaning to the concept of $\sqrt{2}$.

Together, Figures 1 and 2 allow us to pose the question as to whether $\sqrt{2}$ is a rational number. If $\sqrt{2}$ were rational, then it would be possible to draw the square of Figure 2 as a square of circles as in Figure 1, each side having number of circles equal to the numerator of the proposed rational number. That $\sqrt{2}$ is not rational is essentially the same as saying that this cannot be done. Dewdney (1999, pp. 28-29) gives a proof such as the ancient Greeks might have constructed along these lines.

Diagrams can give concrete meaning to concepts which might otherwise remain abstract. Although not always immediately intuitive, like Socrates' guiding of the slave boy, they can be reasoned about until their meaning becomes strikingly clear. Two examples of more interactive diagrams of this nature which one of us has produced are 1. an animation which shows the Theorem of Pythagoras and implicitly its proof (Oldford, 2001) and 2. a three-dimensional physical construction which gives meaning to the statistical concepts of confounding and the role of randomization in establishing causation (Oldford, 1995). In both cases, the visual representation secures the understanding of otherwise abstract concepts.

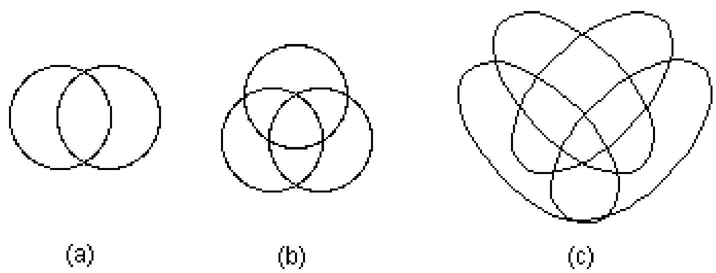

Figure 3: Venn's diagrams.

Venn-like diagrams have a varied history which long predates Venn's use of them (Venn, 1880, 1881). The diagrams have often been given some mystical or religious significance, yet even then the content is conveyed via the same essential features of the diagrams. The overwhelming features of these diagrams are the union and intersection of individual regions.

2.1 The two-ring diagram

Consider diagram (a) of Figure 3. The simple interlocking rings have been used symbolically to represent the intimate union of two as in the marriage of two individuals, or the union of heaven and earth, or of any two worlds (e.g. see Liungman, 1991, Mann 1993). The intersection symbolizes where the two become one. This symbolism is of ancient, possibly prehistoric, origin.

The intersection set, or *vesica piscis* (i.e. fish-shaped container) of Figure 4, has been used by many cultures (the term vesica piscis is also sometimes used for the whole diagram as in Figure 4 (a)). For example, the cover of the famous chalice well at Glastonbury in Somerset England, whose spring waters have been thought of as sacred since earliest times, is decorated with the vesica piscis as in Figure 4 (a). The figure is formed by two circles of equal radius, each having its centre located on the perimeter of the other.

The mystical interpretation might have been amplified by the practical use of the vesica piscis in determining the location and orientation of sacred structures. According to William Stukely's geometric analysis of Stonehenge in 1726, the stones in the inner horseshoe rings seem to be aligned along the curves formed by vesica pisces as in Figure

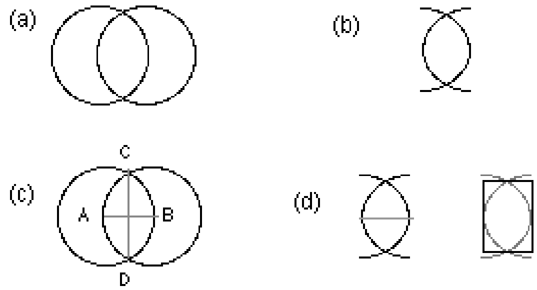

Figure 4: Vesica Piscis.

4(b) (see Mann, 1993, p. 44). Whether Stonehenge's designers had this in mind or not, that Stukely would consider this possibility indicates at least the mystical import accorded the vesica piscis in 1726.

Orientation according to the cardinal axes of the compass were determined via the vesica piscis as follows. The path of the shadow cast by the tip of an upright post or pillar from morning to night determines a west to east line from A to B of Figure 4 (c). The perpendicular line CD is determined by drawing two circles of radius AB, one centred at A, the other at B - a vesica piscis. A rectangular structure with this orientation (or any other significant orientation, e.g. along a sunrise line) and these proportions is easily formed as in Figure 4 (d). Should a square structure be desired (e.g. Hindu temples for the god Purusha, Mann 1993, p. 72) a second vesica piscis can be formed perpendicular to the first (after first drawing a circle of diameter AB centred at the intersection of the lines AB and CD so as to determine a vertical line of length AB to fix the location of the second vesica piscis – the square is then inscribed by the intersection points of the two vesica pisces).

According to Burkhardt (1967, pp. 23-24) (see also Mann, 1993, pp. 71-75) this means of orientation was universal, used in ancient China and Japan and by the ancient Romans to determine the cardinal axes of their cities. The Lady Chapel of Glastonbury Abbey (1184 C.E.) has both its exterior and interior proportions described exactly by rectangles containing a vesica piscis as in Figure 4 (d) (see Mann. 1993, p. 152) and many of the great cathedrals of Europe were oriented using much the same process.

The mathematical structure of the vesica piscis would have been well known and might itself have contributed something to its mystery. The very first geometrical figure appearing in Euclid's Elements is that of Figure 5. Proposition 1 of the first book asserts that an equilateral triangle ABC can be constructed from the line AB, essentially

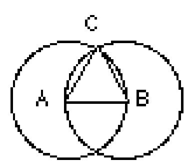

Figure 5: First Figure of Euclid's *Elements*.

by constructing the vesica piscis (see Heath 1908, p. 241).

Interestingly, the equilateral triangle itself has long had a mystical interpretation. According to Liungman (1991), the equilateral triangle is "first and foremost associated with the *holy, divine number of 3*. It is through the tension of opposites that the new is created, the third" (his italics). Xenocrates, a student of Plato, regarded the triangle as a symbol for God. Three appears again in the form of the irrational number $\sqrt{3}$ as the ratio of the length of CD to that of AB in Figure 4 (b). Whether this fact in any way enhanced the mystical significance of the vesica piscis is unknown, although it does seem a plausible speculation – especially for Christian thinkers.

The vesica piscis was adopted as an important symbol in Christianity and appears frequently in Christian art and architecture. Besides the obvious connection with the fish symbol of Figure 4(b) used by early Christians, it came to represent the purity of Christ (possibly through allusion to a stylized womb and so to the virgin birth of Christian scripture). Often the vesica piscis has appeared with a figure of Christ or the Virgin Mary within it (e.g. see Mann, 1993, pp. 24 and 52 for examples from the middle ages). The strength of this symbolism in the Christian faith no doubt significantly contributed to the adoption of the pointed arch (see Figure 6) as a dominant feature in Gothic architecture (e.g. notably in windows and vaults). The vesica piscis continues to be a popular symbol in Christian publications,

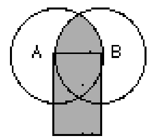

Figure 6: The Gothic arch.

art, and architecture to the present day.

2.2 The three-ring diagram

The three intersecting circles of Venn's diagram in Figure 3(b) is itself an ancient diagram representing a "high spiritual dignity" (Liungman, 1991). As mentioned earlier, the number 3 has long been considered divine. Xenocrates, for example, held the view that human beings had a threefold existence: mind, body, and soul. One can see how, as in the case for two intersecting rings, the union of three different but equal entities each having some attributes in common with another and possibly with all others simultaneously could have a deep mystical or religious appeal.

Certainly, once the holy trinity of the "Father, Son, and Holy Spirit" became established as a fundamental tenet of the Christian faith, the symbols were adopted with the obvious interpretation. The three intersecting rings have long appeared in Christian art and architecture and continue to do so to the present day. Figure 7 shows some variations

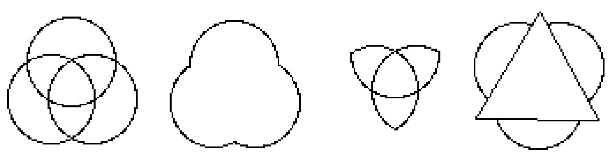

Figure 7: Symbols of the Christian Trinity.

on the three intersecting rings used in Christian symbolism to represent the holy trinity. The last one, interestingly, superimposes the equilateral triangle over the three circles thus making use of two ancient spiritual symbols. This symbol is still commonplace on Christian vestments and altar decorations.

Mathematically, if the circles are drawn (as with the vesica piscis) so that their centres are at the three corners of the intersection set, then the intersection set shares a curious geometric property with a circle – the figure, called a Reuleaux triangle (e.g. see Santalo, 1976, p 8 ff), has constant width through its centre. That is, parallel tangent lines have the same distance between them, wherever they are positioned on the boundary.

2.3 The logic diagrams of Euler

Over the course of one year from 1760 to 1761, the natural scientist and mathematician Leonhard Euler wrote a series of letters to a German princess in which he presented his thoughts on a variety of scientific and philosophical topics with such clarity and generality that the letters were to sweep Europe as "a treasury of science" (Condorcet, p. 12, 1823 preface to Euler) accessible to the reader without much previous knowledge of the subjects addressed.

In the 1823 preface to the third English edition, Euler is regarded as "a philosopher who devote[d] himself to the task of perspicuous illustration." When Euler comes to explain Aristotelian logic to the princess, he makes use of a series of diagrams, diagrams which were to become known in logic as "Eulerian diagrams".

Euler was educated in mathematics as a child by his father, himself a Protestant minister educated in theology and a friend of the great mathematician Johann Bernoulli (e.g. see O'Connor and Robertson, 2001). The plan had been for the younger Euler to study theology at university and this he did, until Bernoulli convinced the father of the young man's formidable mathematical talents. A devout Christian all his life and one-time student of theology, it is hard to imagine that Euler would not have been well aware of the pervasive Christian symbols.

Whatever the source, the diagrams he presented the princess to better explicate Aristotelian logic would be familiar to someone both trained in mathematics and aware of Christian symbolism. The four basic propositions of Aristotle as shown by Euler appear in Figure 8. The diagrams make the points by the intersection (or not) of the circular areas, by

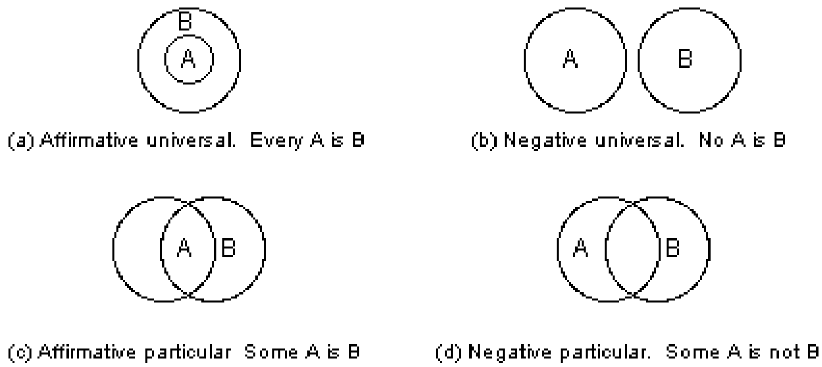

Figure 8: Basic Euler diagrams for the four Aristotelian propositions.

containment (or not) of circular areas, and by containment of the letters A and B – the letter placement allowed Euler to indicate the two "particular" propositions of Figure 8 (c) and (d).

Euler went on to show how all of the Aristotelian syllogisms might be demonstrated in the same way. For example, Figure 9 shows how these diagrams illustrate a relatively simple syllogism.

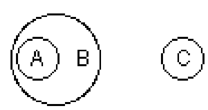

Figure 9: Euler diagram for the syllogism: No B is C; All A is B; ∴ no A is C.

Some syllogisms might need more than one diagram. Figure 10 shows all possible configurations for one such syllogism. Each diagram is itself consistent with the whole of the information contained in the propositions and

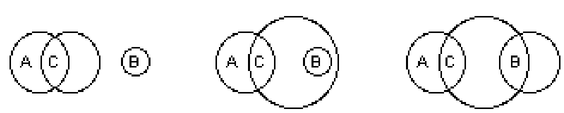

Figure 10: Euler diagrams which are each consistent with the syllogism: No A is B; Some C is A; ∴ some C is not B.

hence in the conclusion of the syllogism. While any one would explain the syllogism, it might be misleading in other respects. Consequently, Euler would completely enumerate the different cases which generate a given syllogism and present them all – nowhere in his letters to the German princess does Euler make use of the three ring diagram of Figure 3(b).

Unfortunately, not all syllogisms can be represented this way. As Venn (1881, pp. 523-4) pointed out even a fairly straightforward proposition such as “All A is either B or C only (i.e. not both)” cannot be expressed with the circles of an Euler diagram. One might attempt to do so via a collection of diagrams as we have done in Figure 11, but individually these do not contain the complete information available in the syllogism and seemingly contradict one another as to what that information might be.

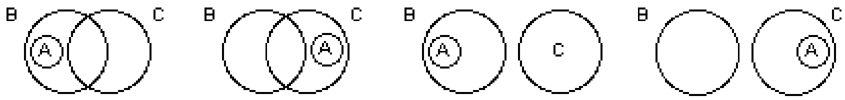

Figure 11: Euler diagrams which collectively express the single proposition: A is either B or C only.

2.4 The logic diagrams of Venn

John Venn graduated from Cambridge University in 1857, was ordained as a Christian priest two years later, and returned to Cambridge in 1862 as a lecturer in “Moral Science” where he studied and taught logic and probability (O’Connor and Robertson, 2001).

Venn was keenly interested in developing a symbolic logic and wanted a diagrammatic representation to go with it. Euler’s diagrams were well known and had widespread appeal by the time of his writing in 1881:

> “Until I came to look somewhat closely into the matter I had not realized how prevalent such an appeal as this had become. Thus of the first sixty logical treatises, published in the last century or so, which were consulted for this purpose:- somewhat at random, as they happened to be most accessible:- it appeared that thirty-four appealed to the aid of diagrams, nearly all making use of the Eulerian Scheme.”
> John Venn, *Symbolic Logic*, 1881 (page 110 of the 2nd Edition, 1894).

Venn’s logic, like Boole’s, was mathematical in nature. For example, $xy\bar{z} = 0$ indicates that the simultaneous condition x and y and not z cannot occur. The mathematics allowed propositions such as this to accumulate and inferences to be drawn as the information became available. Venn’s diagrams had to serve in the same way. In his words:

> “Of course we must positively insist that our diagrammatic scheme and our purely symbolic scheme shall be in complete correspondence and harmony with each other. *The main objection of the common or Eulerian diagrams is that such correspondence is not secured.* ... But symbolic and diagrammatic systems are to some extent artificial, and *they ought therefore to be so constructed as to work in perfect harmony together*.”
> John Venn, *Symbolic Logic*, 1881 (page 139 of the 2nd Edition, 1894).
> Italic emphasis is added.

Besides the failings alluded to in the previous section, Euler’s diagrams required considerable thought in the construction – all possibilities needed to be followed as the diagrams were constructed. If you know the answer, as is the case for simple syllogisms, the diagrams are easy to construct; if you don’t they can be considerable work.

Euler diagrams were designed to *demonstrate* the known content of a syllogism; Venn’s diagrams were designed to *derive* the content. Remarkably, this profound distinction between the two diagrams can be missed by some mathematical popularizers, notably Dunham (1994 p. 262) who imagines Venn’s innovation being discovered by any “child with a crayon”.

Given his religious training, it would be surprising if Venn were unaware of the Christian symbolism of at least the three ring diagram he was to introduce to the study of logic. This three-ring diagram was to be employed to record the logical content of each proposition *as it became available*.

Figure 12 illustrates this use for a simple syllogism – one shades out the regions which correspond to impossible conditions as they become known. In this way, information accumulates by being added to the diagram as it becomes available. At any point one can see the consequences of the information to date – only the unshaded regions (including the region outside all three circles: not A not B not C) are possible.

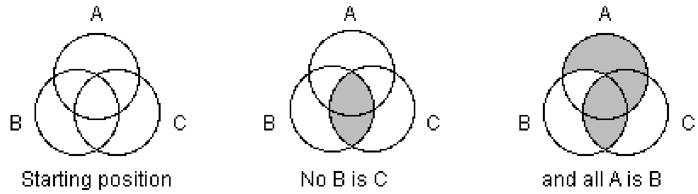

Figure 12: No B is C; All A is B; therefore no A is C.

Figure 13 illustrates a more complicated syllogism which requires Venn's diagram of Figure 3(c) (which seems to be original to Venn) in order to render the logic diagrammatically. Left to right the diagrams show the effect of adding

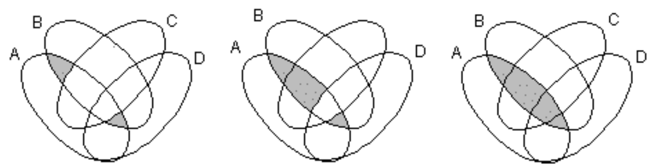

Figure 13: A complex syllogism – the information of each statement is added to the diagram by progressively shading those regions which the statement excludes. From left to right the cumulative effect of the following statements can be read from the diagrams: i. All A is either B and C, or not B; ii. If any A and B is C, then it is D; and iii. No A and D is B and C. From the last figure we see that together these statements imply that no A is B.

each new piece of information to what is known. Carrying out the logic via Euler diagrams would be considerably more difficult.

Besides their *active* use in the analysis of logical structure, Venn's diagrams differ from Euler's in another important respect. Each region represents a class; unshaded it remains possible, shaded it becomes impossible. There is no provision for indicating the particular "Some A is B" – it remains indistinguishable from "A and B has not been ruled out". Venn sees no need to explicitly distinguish these possibilities; they remain only because of the historical dominance of Aristotelian logic.

2.5 The essence of Venn diagrams

Throughout their long history, Venn-like diagrams seem to be put to similar use, albeit in different contexts. The diagrams compel one to think in terms of identifying different entities, what they have in common, and how they differ from one another and possibly from everything else. As formal set theory developed, the same figures were used to naturally embody the properties of sets – intersection, union, complement. However, just as some ideas can be given meaning only by a diagram, a diagram can be incapable of easily producing anything but these ideas.

3 Weakness of Venn Diagrams for Teaching Probability

Venn diagrams, as an extension of Euler diagrams, are a useful tool in *logic* where conditions are either possible or impossible. Because the rules of probability are based on events and because events are traditionally represented as sets, Venn diagrams would seem well suited also for illustrating probability concepts; this is not the case for three main reasons.

3.1 Teaching logic under the guise of probability

Venn diagrams skew the teaching of introductory probability towards what are fundamentally problems in logic which only involve probability incidentally because the entities being manipulated happen to be probabilities.

For instance, a basic relationship in symbolic logic, self evident in set theory via Venn diagrams, is $A \cup B = A + B - A \cap B$. Introductory probability texts merely exercise this idea, and its extension to three events, in its probability version $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$ (which says little more than probability behaves as a measure on sets). Two typical examples are:

> Illustration 1: *Paul and Sarah both apply for jobs at a local shopping centre; the probability Paul gets a job is 0.4, the probability Sarah gets a job is 0.45 and the probability they both get jobs is 0.1.*
> *What is the probability at least one of them is employed?*
>
> Illustration 2: *Suppose that 75% of all homeowners fertilize their lawns, 60% apply herbicides and 35% apply insecticides. In addition, suppose that 20% apply none of these, 30% apply all three, 56% apply herbicides and fertilizer, and 33% apply insecticides and fertilizer.*
> *What percentage apply (a) herbicides and insecticides; (b) herbicides and insecticides but not fertilizer?*

While the training in logic that such problems provide may be useful, this is outweighed by several disadvantages:

1. The values given for the probabilities in such problems would, in practice, have to come from survey data (e.g. about employment success or lawncare practices). The probabilites asked for in the questions would then exist as relative frequencies in the data; it is only the artificial *selective* revelation of data characteristics (proportions) that allow the problem to be posed as an exercise in logic.

2. The probabilities of 0.4 and 0.45 for Paul and Sarah are misleading – what data would provide is such a probability for a randomly-selected person with particular characteristics of sex, age, etc. .

3. It is unclear in Illustration 1 what data would yield the estimate of the joint probability of 0.1.

4. It is unclear why any of the probabilities asked for is of interest; except as an exercise in logic, the student could regard these as mere 'make-work' problems.

The first three disadvantages are already leading the student in unprofitable directions with regard to the use of probabilistic ideas in statistics; this latter is the reason most students study introductory probability. The last disadvantage tends to trivialize a field whose proper study is important, both for its use in statistics and in its own right.

Such problems are typically artificial and give no insight into probability beyond the mathematical manipulation of sets. It is only once some axioms for probability are in place that we have the corresponding probability results. The Venn diagrams of Figure 3 give such prominence to the inclusion exclusion principle that it is commonplace for introductory treatments of probability to fall into the trap of framing probability problems just to exercise this principle.

3.2 Confusing the nature of relationships

A key idea in probability is independence as one pole in describing the continuum of relationships. At best, Venn diagrams convey little information about independence and, when the idea of disjointness is included, they can be actively misleading.

For example, a viewer of Figure 14 could be forgiven for thinking, wrongly, that the respective diagrams (a) and (b) represent independent and dependent events. Figure 14 (a) shows $Pr(A|B) = Pr(B|A) = 0$ – that is, *if one event occurs, the other cannot*. Thus, except for A or B being impossible events, $Pr(A|B) \neq Pr(A)$ and $Pr(B|A) \neq Pr(B)$ so that, despite the clear visual suggestion to the contrary, the events are dependent. For Figure 14 (b), probabilities can be associated with events A and B so that they are independent, again contrary to what is suggested visually. It should therefore not surprise us when students confuse disjoint events with independent events when these ideas are introduced using Venn diagrams.

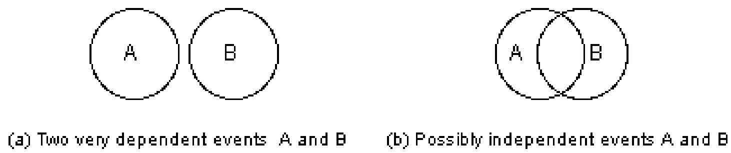

Figure 14: Counter intuitive diagrams for probability.

3.3 Inability to quantify probabilities

The inherent ability of Venn diagrams to distinguish the dichotomy of what is possible and what is impossible does not lend itself to quantifying probabilities on a continuous scale. This is obviously the case with the roughly circular or elliptical shapes commonly used for events within a sample space, and the situation is not greatly improved if squares or rectangles are used instead.

However, just as Venn's apparently minor adaptation of Euler diagrams substantially enhanced their usefulness in logic, so what we call an eikosogram takes the idea of a Venn diagram with rectangular areas and adapts it to provide a powerful tool for visualizing probabilities. The rectangular shapes of the regions of an eikosogram provide a natural scale for quantifying probabilities and their layout gives appropriate visual emphasis to the regions/events.

4 Eikosograms

For probability, Venn's diagrams fall far short of satisfying his own dictum (Venn, 1881, p.139) that "…symbolic and diagrammatic systems …be so constructed as to work in perfect harmony together." – no surprise since they were designed for Venn's symbolic logic system, not for probability. A diagram tailored to probability and one which arguably fulfills Wittgenstein's notion of an "ostensive definition" for probability (especially for conditional probability) is the *eikosogram* – a word[2] constructed to evoke 'probability picture' from classical Greek words for probability (*eikos*) and drawing or writing (*gramma*).

Just as the ring diagrams were not new to Venn, so too this diagram has seen use before – variants of it have been used to describe observed frequencies for centuries (at least as early as 1693 by Halley; see Friendly 2002 for some history on these variants). Recently Michael Friendly has developed and promoted a variant he calls "mosaic plots" to display observed frequencies upon which fitted model residuals are layered using colour (Friendly, 1994). The earliest use of an eikosogram (i.e. displaying probabilities) of which we are aware is by Edwards (1972, p. 47) where a single diagram appears with the unfortunate label of 'Venn diagram'. The label is an example of how far the sense of a Venn diagram has been stretched.

Certainly teachers of probability have long used relative areas when teaching probability. That such diagrams have been used and developed independently by many authors over time speaks to their naturalness and consequent value in describing and understanding probability.

All eikosgrams are built on a unit square whose unit area represents the probability 1, or certainty. An eikosogram is constructed by dividing the square first into vertical strips, each one corresponding to a conditioning event and the width determined by that event's probability. Each strip is then divided horizontally according to the values of probabilities conditional on the event defining the vertical strip. All resulting rectangular blocks have areas equal to the probabilities involved. Shading is used to distinguish the blocks vertically. This definition will become clear with a few eikosograms.

4.1 The basic eikosogram.

The basic eikosogram is that of a single event for which no conditioning event is considered and so no division into vertical strips is made. Suppose we have such an event, A say, which occurs with probability of $1/3$. Then the eikosogram representing this probability is shown in Figure 15.

[2]This construction was kindly suggested by our colleague Prof. G.W. Bennett.

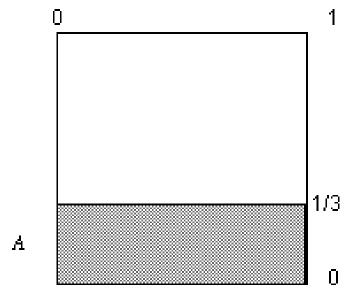

Figure 15: Eikosogram: Shows $Pr(A) = 1/3$.

The unit square is divided only horizontally at $1/3$, and the area of the shaded region gives the probability of the event A occurring. Horizontal and vertical positions can be read off the top and right sides of the unit square, so rectangular areas are easily calculated (having these sides of the square as labelled axes produces a left to right order in reading the diagram as in the symbolic statement $Pr(A) = 1/3$). The unshaded region has area equal to the probability that the event A does not occur (here $2/3$, from $1 - 1/3$).

A physical analogy to give meaning to the probability is easily had. Imagine this eikosogram lying flat on the ground in the rain; of those raindrops which hit the square, the proportion which strike the shaded region corresponds to the probability that the event A occurs. This could be easily simulated by Monte Carlo and displayed on a computer screen (cf Oldford, 2001b).

All characteristics of this simple picture are true to the idea of probability; none is misleading. Already, the following points can be made:

- The idea of “odds” follows by pointing out that twice as many of the raindrops striking the square will miss the shaded region as will hit it. We say that the odds are 2 to 1 against A (or 1 to 2 in favour of A), as determined by the ratio of the relevant areas.
- Because all probabilities are areas within (or equal to) the unit square, the diagram shows that probabilities can only take on real values from 0 to 1 inclusive.
- The probabilities which correspond to A occurring and A not occurring must sum to one because their regions clearly divide the unit square; symbolically we have $Pr(A\ not\ occurring) = 1 - Pr(A\ occurring)$.
- More generally, the areas of non-overlapping regions which cover the unit square sum to one.

Axioms for probability are naturally embedded in this picture. Note that the complement of A, a set theoretic term, is unnecessary at this point and should be avoided; that the event A either occurs or does not (i.e. a raindrop strikes the shaded area or it does not) is quite natural and appears as such in the eikosogram.

To capitalize on this, one could introduce a random variable, say Y, which takes one of two values to indicate whether A occurs or does not. If A occurs, Y has value “`y`” (short for “A occurs”); if A does not occur, then Y takes value “`n`” (short for “A does not occur”). The eikosogram of Figure 15 could then be redisplayed using Y as in Figure 16. In any application, the variate Y and its values will be more meaningful. For example, Y might represent gender and hence take values of “`male`” and “`female`” rather than “`n`” and “`y`” resulting in a more meaningful labelling of the eikosogram. Examples abound and could easily be constructed in class.

From Figure 16 we can read directly that $Pr(Y = \mathtt{y}) = 1/3$ and that $Pr(Y = \mathtt{n}) = 2/3$ Together these two numbers determine what is called the probability distribution of the binary random variable Y, denoted by $Pr(Y)$. Each such distribution will produce its own eikosogram; the eikosogram is 1-1 with the distribution. It is a short step from the eikosogram to the more traditional display of this distribution as shown in Figure 17.

This bar-chart is well-suited to display the characteristics of the distribution of Y, not least because the bar heights share a common vertical axis, the elementary graphical perceptual task at which humans excel (e.g., see p. 254 of Cleveland, 1985). For either assessment or comparisons of distributions, particularly if either Y takes on many values or the values Y takes can be meaningfully ordered along its horizontal axis, this diagram will be superior to the

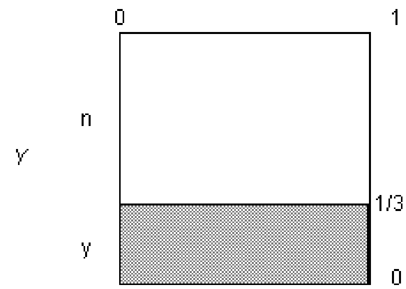

Figure 16: Eikosogram: Shows $Pr(Y = \mathtt{y}) = 1/3$.

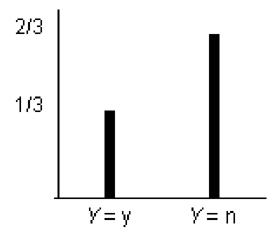

Figure 17: Distribution for the random variable Y.

eikosogram. The superiority of the eikosogram lies rather in the development of an understanding of probability and its rules, something which must precede the comparison of whole distributions.

4.2 Conditional and joint probabilities

The explanatory power of the eikosogram is put to fuller use when more than one random variable is considered. In this case, the ideas of conditional and joint probabilities arise in addition to the marginal ones.

Conditional probability is introduced to the student by showing them the eikosogram of Figure 18. There we see

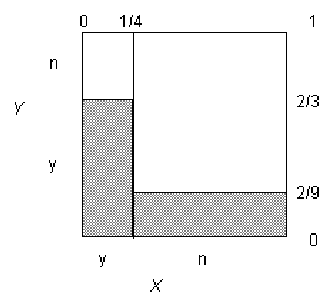

Figure 18: Eikosogram for Y given X.

that a second variable X has been introduced which like Y takes on two values $X = \mathtt{y}$ (the left vertical strip) and $X = \mathtt{n}$ (the right vertical strip). As before the shaded area corresponds to the probability that $Y = \mathtt{y}$, the unshaded to $Y = \mathtt{n}$.

Again, the raindrop metaphor can be put to good use in giving a direct interpretation of the various probabilities involved. The probability of any event is the area of that region of the unit square matching the event.

From Figure 18, the region corresponding to the event $X = \mathtt{y}$ is the entire left vertical strip. From the diagram, this rectangular area is simply the $width \times height = 1/4 \times 1 = 1/4$, so $Pr(X = \mathtt{y}) = 1/4$. Similarly $Pr(X = \mathtt{n}) = 3/4 = 1 - Pr(X = \mathtt{y})$. In the case of vertical strips, the probabilities can be determined directly from the horizontal axis at the top of the eikosogram since each entire vertical strip will have $height = 1$ (i.e. these marginal probabilities determine the width of the strips).

Determining $Pr(Y = \mathtt{y})$ amounts to summing the areas of the two shaded rectangles, which from Figure 18 is easily seen to be $1/4 \times 2/3 + 3/4 \times 2/9 = 1/3$. Figure 18 was constructed with probabilities to match those in Figure 16; Figure 16 is the display of the marginal distribution of Y corresponding to the joint of X and Y seen in Figure 18.

One way of imagining this derivation of the marginal distribution of Y is to think of the eikosogram of Figure 18 as a water container with the shaded areas corresponding to the level of water in each of two separate chambers: one being the left vertical strip with water filling 2/3 of the chamber, the other being the right vertical strip with water filling only 2/9 of this chamber. Imagine further that the line making the vertical division at 1/4 is actually a removable barrier which has created the separate chambers. Finding the marginal distribution of Y amounts to removing this barrier and having the water settle to some new level in the whole container as seen in Figure 16.

Conditional probability is introduced via Figure18 by considering each vertical strip in turn. The leftmost strip fixes the condition $X = \mathtt{y}$. When we ask the question ‘Of those raindrops which strike the leftmost strip, what proportion lands on the shaded area?’, then we are asking for the probability that $Y = \mathtt{y}$ conditional on, or given that, $X = \mathtt{y}$, or symbolically for $Pr(Y = \mathtt{y}|X = \mathtt{y})$. The raindrop metaphor makes it clear that this conditional probability is the ratio of the area of the bottom left shaded rectangle to the area of the rectangle which is the leftmost strip.

Since the width of both of these rectangles is identical by design (1/4 in Figure 18), this amounts to asking for the relative height of the smaller shaded one to the larger rectangular strip which contains it. This in turn amounts to asking for the absolute height of the bottom left shaded rectangle (since again by design, the strip’s height is 1). Reading from the vertical axis of Figure 18 we see that the desired value is 2/3. While we could have similarly found that $Pr(Y = \mathtt{n}|X = \mathtt{y}) = 1/3$ it is apparent from the diagram that symbolically we must have $Pr(Y = \mathtt{n}|X = \mathtt{y}) = 1 - Pr(Y = \mathtt{y}|X = \mathtt{y})$ and so the value 1/3. Note that the point could now be made that although $Pr(Y = \mathtt{y}|X = \mathtt{y})$ and $Pr(Y = \mathtt{n}|X = \mathtt{y})$ must sum to one, $Pr(Y = \mathtt{y}|X = \mathtt{y})$ and $Pr(Y = \mathtt{y}|X = \mathtt{n})$ need not (a conceptual mistake sometimes by students).

It is as if we isolated the leftmost strip, widened it to width 1, and read off the vertical value of a basic eikosogram like that of Figure 16 except having shaded height of 2/3. The leftmost strip (widened to have width 1) displays the conditional probability distribution for Y given $X = \mathtt{y}$. To emphasize the point, simply draw the corresponding basic eikosogram when $X = \mathtt{y}$. If all individual eikosograms for every vertical strip are imagined drawn separately, it becomes apparent that the joint distribution can be thought of as the weighted collection of conditional distributions, where the weights given by the marginal probabilities of each strip (here 1/4 for $X = \mathtt{y}$ and 3/4 for $X = \mathtt{n}$) are identified with the widths for the eikosogram of the joint distribution. The joint is thus shown to be a *mixture* of the conditionals, formed by pushing together the individual (i.e. conditional) eikosograms having the correct width. In this way complex eikosograms can be built up from simpler ones and conversely simpler ones had from complex ones.

4.2.1 Probability calculation rules.

When we calculate areas on the eikosogram of Figure 18, all essential relationships between probabilities tumble out.[3]

Once the conditional probabilities just determined are understood, then rules for calculating joint probabilities from marginal and conditional can be introduced by simply calculating the corresponding areas. From Figure 18 these are demonstrably as follows:

$$\begin{aligned} Pr(Y = \mathtt{y} \text{ and } X = \mathtt{y}) &= Pr(Y = \mathtt{y}|X = \mathtt{y}) \times Pr(X = \mathtt{y}) \\ &= 2/3 \times 1/4 = 1/6 \end{aligned}$$

$$\begin{aligned} Pr(Y = \mathtt{n} \text{ and } X = \mathtt{y}) &= Pr(Y = \mathtt{n}|X = \mathtt{y}) \times Pr(X = \mathtt{y}) \\ &= 1/3 \times 1/4 = 1/12 \end{aligned}$$

[3]All of these results hold for eikosograms with any finite number of values for X or for Y or for both X and Y. Neither need be only binary. Formally, for infinitely many values, some extension would be required as the eikosograms would not be defined. The move to probability density functions would be opportune then, perhaps following a transition much like that from Figure 16 to Figure 17.

$$\begin{aligned} Pr(Y = \mathtt{y} \text{ and } X = \mathtt{n}) &= Pr(Y = \mathtt{y}|X = \mathtt{n}) \times Pr(X = \mathtt{n}) \\ &= 2/9 \times 3/4 = 1/6 \end{aligned}$$

$$\begin{aligned} Pr(Y = \mathtt{n} \text{ and } X = \mathtt{n}) &= Pr(Y = \mathtt{n}|X = \mathtt{n}) \times Pr(X = \mathtt{n}) \\ &= 7/9 \times 3/4 = 7/12 \end{aligned}$$

which of course sum to 1. Together these values determine what is called the joint probability distribution of X and Y and is generally written as $Pr(X \text{ and } Y)$ or more compactly as as $Pr(X, Y)$. The general calculation rule used here was that of the $Area(rectangle) = width \times height$ and applied whatever the value of X or Y. The corresponding rule of probability is therefore expressed as:

$$Pr(X, Y) = Pr(Y|X) \times Pr(X)$$

Rules for calculating marginal probabilities from joint are easily demonstrated from Figure 18 by determining $Pr(Y = \mathtt{y})$. This probability must be the total area of the shaded regions corresponding to the event $Y = \mathtt{y}$. Mathematically, one sees immediately that marginal probabilities are determined by summing over the relevant pieces of the joint distribution as in

$$\begin{aligned} Pr(Y = \mathtt{y}) &= Pr(Y = \mathtt{y} \text{ and } X = \mathtt{y}) + Pr(Y = \mathtt{y} \text{ and } X = \mathtt{n}) \\ &= 1/6 + 1/6 \;\; = \;\; 1/3 \\ &= 1 - Pr(Y = \mathtt{n}). \end{aligned}$$

Bayes' rule follows directly from calculating the only remaining probabilities, namely the conditional probability of $X = \mathtt{y}$ or $X = \mathtt{n}$ given $Y = \mathtt{y}$ or $Y = \mathtt{n}$. Conditioning on $Y = \mathtt{y}$ amounts to considering only the shaded regions of Figure 18. We are asking of those raindrops which strike a shaded area, what proportion also fall on the leftmost strip where $X = \mathtt{y}$? Finding the $Pr(X = \mathtt{y}|Y = \mathtt{y})$, say, is equivalent to finding the ratio of the leftmost shaded area to the total shaded area.

Bayes' rule falls out as a consequence:

$$\begin{aligned} Pr(X = \mathtt{y}|Y = \mathtt{y}) &= Pr(Y = \mathtt{y}|X = \mathtt{y})Pr(X = \mathtt{y})/Pr(Y = \mathtt{y}) \\ &= (1/2) * (1/3)/(7/18) \;\; = \;\; 3/7, \end{aligned}$$

or equivalently

$$\begin{aligned} Pr(X = \mathtt{y}|Y = \mathtt{y}) &= Pr(Y = \mathtt{y} \text{ and } X = \mathtt{y})/Pr(Y = \mathtt{y}) \\ &= (1/6)/(7/18) \;\; = \;\; 3/7. \end{aligned}$$

The general Bayes' rule is expressed as either

$$Pr(X|Y) = Pr(Y|X) \times Pr(X)/Pr(Y)$$

or more compactly as

$$Pr(X|Y) = Pr(X, Y)/Pr(Y)$$

Had we drawn the probability strips by conditioning on $Y = \mathtt{y}$ and $Y = \mathtt{n}$, rather than $X = \mathtt{y}$ and $X = \mathtt{n}$, then the eikosogram would appear as in Figure 19. Note that the events have the same areas as before. Transforming the eikosogram from that of Figure 18 to that of Figure 19 is a good exercise in probability calculation for the student. It requires determining first one of $Pr(Y = \mathtt{y})$ or $Pr(Y = \mathtt{n})$ to fix the location of the vertical strip, then each of the conditional probabilities $Pr(X = \mathtt{y}|Y = \mathtt{y})$ and $Pr(X = \mathtt{y}|Y = \mathtt{n})$ to determine the heights of each shaded rectangle.

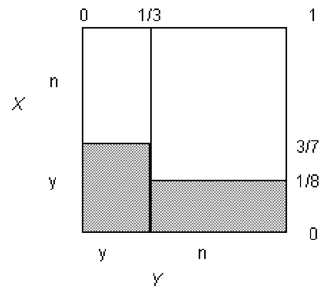

Figure 19: Eikosogram for X given Y. This is one to one with the eikosogram for Y given X given in Figure 18

4.3 Probabilistic independence

Probabilistic independence is a much more subtle concept than most introductory treatments of probability would have one believe. In particular, independence of events can and should be carefully and explicitly distinguished from independent random variables, yet this is rarely the case. Whereas Venn diagrams are ill-suited to, and even misleading for, elucidating the probabilistic independence of events, they are quite incapable of distinguishing independent events from independent random variables. Eikosograms on the other hand, seem well suited to exploring independence.

Consider again the eikosogram of Figure 18 from which it can be seen that

$$Pr(Y = \mathrm{y}) \neq Pr(Y = \mathrm{y}|X = \mathrm{y})$$

The left hand side of the equation is the proportion of raindrops which strike the shaded area (i.e. $Y = \mathrm{y}$) of Figure 18. The right side of the equation, on the other hand, restricts focus to those raindrops striking the leftmost strip of Figure 18 (i.e. $X = \mathrm{y}$) and gives the proportion of these which strike a shaded area. The inequality states simply that the proportion of raindrops striking a shaded area *depends* on whether you are considering the figure as a whole or just the one strip.

Formally we say that the event $Y = \mathrm{y}$ *depends* on the event $X = \mathrm{y}$. It can be determined that we also have that the event $X = \mathrm{y}$ depends on the event $Y = \mathrm{y}$ (either directly from the eikosogram, or formally as derivation using the calculation rules for probability). This symmetry always holds. Consequently, we talk about the events $Y = \mathrm{y}$ and $X = \mathrm{y}$ symmetrically as being *dependent events*.

If instead we have

$$Pr(Y = \mathrm{y}) = Pr(Y = \mathrm{y}|X = \mathrm{y})$$

then the proportion of raindrops striking a shaded area is the same whether we consider just the one strip, or the figure as a whole. We say that the event $Y = \mathrm{y}$ *does not depend on*, or is *independent* from, the event $X = \mathrm{y}$. More symmetrically, we say that the events $Y = \mathrm{y}$ and $X = \mathrm{y}$ are *independent events*. Figure 20 shows the eikosogram for

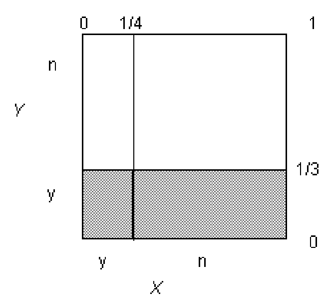

Figure 20: Independent events from independent random variables X and Y.

which this is the case (and $Pr(Y = \mathtt{y}) = 1/3$ to be consistent with Figure 16).

The striking characteristic of this eikosogram is that it is flat – the shaded areas have the same vertical coordinate, in this case 1/3. If the vertical line at 1/4 were removed as well as any reference to X and the values it can take, then Figure 20 would be identical to Figure 16. In terms of the water container metaphor, removing the vertical barrier at 1/4 has no effect on the water levels in either container. This flatness (or common water level) is an essential characteristic of probabilistic independence in an eikosogram.

This flatness also indicates that in addition to independent events $Y = \mathtt{y}$ and $X = \mathtt{y}$, we also have independence of the events $Y = \mathtt{n}$ and $X = \mathtt{n}$, of the events $Y = \mathtt{y}$ and $X = \mathtt{n}$, and of the events $Y = \mathtt{n}$ and $X = \mathtt{y}$. That is, the independence holds for all possible values of the variables Y and X. When this is the case, we say that Y and X are *independent random variables* and express this symbolically either as

$$Pr(Y) = Pr(Y|X)$$

or equivalently as

$$Pr(X) = Pr(X|Y)$$

either of which imply via the (rectangle area) calculation rule that

$$Pr(X, Y) = Pr(X) \times Pr(Y).$$

This last expression (or the corresponding one for events) is sometimes taken ab initio to define probabilistic independence, a choice which can appear to be arbitrary. The route just taken through conditional probability, which instead derives this multiplicative rule for independence, seems more natural and compelling.

Symbolically we denote independence with a '$\perp\!\!\!\perp$' as in $Y \perp\!\!\!\perp X$ for the independence of the random variables and $(Y = \mathtt{y}) \perp\!\!\!\perp (X = \mathtt{y})$ for the events. Dependence will be indicated using the same symbol but with a stroke through it as in $Y \not\perp\!\!\!\perp X$ when Y and X are known to be dependent (similarly for events).

In this example, the flatness indicated independence both of the events $Y = \mathtt{y}$ and $X = \mathtt{y}$ and of the random variables Y and X. Figure 21 shows a case where if X takes on more than two values, say $X = \mathtt{a}$, $X = \mathtt{b}$, or $X = \mathtt{c}$,

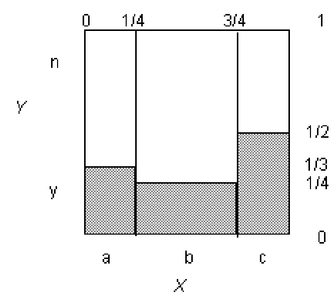

Figure 21: Dependent random variables Y and X. Independent events $Y = \mathtt{y}$ and $X = \mathtt{a}$ since $Pr(Y = \mathtt{y}|X = \mathtt{a}) = Pr(Y = \mathtt{y})$.

then we can have independent events $Y = \mathtt{y}$ and $X = \mathtt{a}$ but *dependent random variables* Y and X. Symbolically we can have $(Y = \mathtt{y}) \perp\!\!\!\perp (X = \mathtt{a})$ yet $Y \not\perp\!\!\!\perp X$.

The independence of the two events can be determined in any one of several ways:

- The appropriate calculation could be done directly from the eikosogram of Figure 21 by calculating the sum of all shaded areas and observing this to be equal to the height of the leftmost shaded bar, namely 1/3.

- The eikosogram could be transformed to one which considers only the cases in which the events of interest either occur or do not occur. For X this amounts to the cases $X = \mathtt{a}$ and $X \neq \mathtt{a}$ which is to say *either* $X = \mathtt{b}$ *or* $X = \mathtt{c}$.

  The eikosogram for this is had from Figure 21 by removing the vertical barrier at 3/4 and allowing the water of the two rightmost containers to mix and settle at a common level. The common level would be 1/3 and the resulting eikosogram would be identical to that of Figure 20 except that instead of "$X = \texttt{y}$" and "$X = \texttt{n}$" we would have "$X = \texttt{a}$" and "$X = \texttt{b}\ or\ \texttt{c}$". The flatness would allow us to immediately conclude the independence of the events.

- If the eikosogram of Figure 16 is available, then simply noticing that the height of the shaded bar there (i.e the unconditional probability) is identical to that of the leftmost shaded bar in Figure 21 (the conditional probability) is sufficient to declare the independence of the events $Y = \texttt{y}$ and $X = \texttt{a}$.

Each of these approaches provides the student with different insights into the nature of independence.

The dependence of the random variables is indicated from the eikosogram by the varying heights of the shaded bars; had these all been the same height (whatever the widths) the variables would have been independent. The flatness of the eikosogram for two random variables is both necessary and sufficient for independence of the variables.

Independence of events is easily seen to be a special case of independence of random variables. As in the second bullet above, we can see that the independence of events looks for flatness in an eikosogram involving only binary random variables indicating the occurrence, or not, of the events in question. Flatness here is coincident with the independence of these two binary random variables, which in turn is coincident with the independence of the events.

A random variable is a broad concept, one which is used to label a collection of mutually exclusive events (e.g. X covers each of the events $X = \texttt{a}$, $X = \texttt{b}$, or $X = \texttt{c}$). The independence of two random variables is thus seen to be a broad assertion about the independence of many different events. While it is the case that $Y \perp\!\!\!\perp X \Rightarrow (Y = \texttt{y}) \perp\!\!\!\perp (X = \texttt{a})$ the above example shows that the converse is not true.

4.4 Conditional independence

Once probabilistic independence has been explored with two random variables, conditional independence (dependence) can be introduced. Because events are always to be distinguished from variables, the simplest way to proceed is with three binary variables X, Y, and Z whose discussion will cover both cases.

Figure 22 gives an eikosogram which illustrates many of the concepts (N.B. this eikosogram has *not* been con-

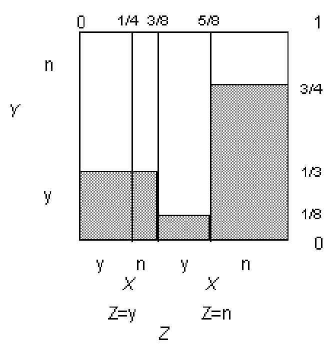

Figure 22: Random variables Y and X are conditionally independent given $Z = \texttt{y}$ but are *not* conditionally independent given $Z = \texttt{n}$. Symbolically $Y \perp\!\!\!\perp X | (Z = \texttt{y})$ but $Y \not\perp\!\!\!\perp X | (Z = \texttt{n})$

structed to agree with that of Figure 16, i.e. $Pr(Y = \texttt{y}) \neq 1/3$). As before, the conditioning variable values (or events) are given along the horizontal axis. With three variables there are six different eikosograms possible: one of three variables must be placed on the vertical axis and for each of these the two horizontal variables could be interchanged. In practice, it is the exchange of variables on the vertical axis which matters most.

This eikosogram is interpreted in a fashion similar to that for two variables. One can essentially read off

- the joint probabilities for all combinations of X and Z
(e.g. $Pr(X = \texttt{y}\ and\ Z = \texttt{y}) = 1/4$, $Pr(X = \texttt{n}\ and\ Z = \texttt{y}) = 3/8 - 1/4 = 1/8$, etc.),

- the marginal probabilities of Z
(i.e. $Pr(Z = \texttt{y}) = 3/8$ and $Pr(Z = \texttt{n}) = 1 - 3/8 = 5/8$),

- the marginal probabilities of X
(i.e. $Pr(X = \mathtt{y}) = 1/4 + (5/8 - 3/8) = 1/2$ and $Pr(X = \mathtt{n}) = 1 - 1/2 = 1/2$),

- and easiest of all the conditional probabilities of Y given each pair of values for X and Z
(e.g. $Pr(Y = \mathtt{y}|Z = \mathtt{n}\, and\, X = \mathtt{y}) = 1/8$).

Other probabilities require a little more calculation. For example $Pr(Y = \mathtt{y})$ is the sum of all shaded areas and $Pr(Y = \mathtt{y}|X = \mathtt{y})$ is the proportion of the area in the vertical strips having $X = \mathtt{y}$ that is shaded. Calculating other joint or conditional probabilities amounts to similar calculations of the relevant rectangular areas.

The flat area at the left of this eikosogram is indicative of some sort of independence when $Z = \mathtt{y}$. In particular, it implies the independence of the random variables Y and X *provided* $Z = \mathtt{y}$. We say that the *random variables* Y and X are *conditionally independent* given the *event* $Z = \mathtt{y}$ and express this symbolically as $Y \perp\!\!\!\perp X|(Z = \mathtt{y})$. Similarly we can see that the events $Y = \mathtt{y}$ and $X = \mathtt{y}$ are conditionally independent given $Z = \mathtt{y}$, or symbolically $(Y = \mathtt{y}) \perp\!\!\!\perp (X = \mathtt{y})|(Z = \mathtt{y})$. Other events associated with this flat area are conditionally independent given the event $Z = \mathtt{y}$.

Conditional independence occurs when shaded bars in an eikosogram have the same height (to make a contiguous flat area requires only rearrangement of the conditioning events along the horizontal axis). No flat area on the right (i.e. $Z = \mathtt{n}$) of the eikosogram of Figure 22 means these independencies do not hold when $Z = \mathtt{n}$. That is $Y \not\perp\!\!\!\perp X|(Z = \mathtt{n})$ and $(Y = \mathtt{y}) \not\perp\!\!\!\perp (X = \mathtt{y})|(Z = \mathtt{n})$. Had the area at the left not been flat, then the conditional independence there would have disappeared as well.

Figure 23 is similar to Figure 22 matching all of its probabilities but the conditional probabilities of Y given X

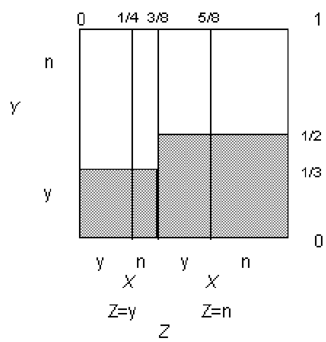

Figure 23: Random variables Y and X are conditionally independent given Z. Symbolically, $Y \perp\!\!\!\perp X|Z$.

when $Z = \mathtt{n}$. In this configuration there are flats both when $Z = \mathtt{y}$ and when $Z = \mathtt{n}$. Because there is a flat for each value of Z, we say that the random variables Y and X are conditionally independent given Z and write $Y \perp\!\!\!\perp X|Z$. It is clear both notationally and from the comparison of Figures 22 and 23 that $Y \perp\!\!\!\perp X|Z$ is a much stronger condition than $Y \perp\!\!\!\perp X|(Z = \mathtt{y})$.[4]

Were the flats all to occur at the same level, as in Figure 24, then more independencies must hold. In particular all of the following hold iff there is a single flat: conditionally $Y \perp\!\!\!\perp X|Z$ and $Y \perp\!\!\!\perp Z|X$; and unconditionally $Y \perp\!\!\!\perp X$, and $Y \perp\!\!\!\perp Z$.

The flat says nothing about the relationship between the conditioning variables X and Z. In this figure they are dependent both unconditionally and given Y. This can be seen by the fact that the ratio of the width of strip $X = \mathtt{y}$ to that of the strip $X = \mathtt{n}$ is different depending on whether $Z = \mathtt{y}$ or $Z = \mathtt{n}$. The ratio when $Z = \mathtt{y}$ is 1/4:1/8 or 2:1 and when $Z = \mathtt{n}$ it is 1/4:3/8 or 2:3. These correspond to the odds of $X = \mathtt{y}$ to $X = \mathtt{n}$ when $Z = \mathtt{y}$ and when $Z = \mathtt{n}$, respectively. Had they been equal, then we would have had $X \perp\!\!\!\perp Z$.

Had the ratios been the same, then this together with the flat constitute necessary and sufficient conditions for the *mutual independence* of all three variables X, Y, and Z. An example of such an eikosogram is given in Figure 25.

[4]For a variety of reasons (not least of which is model simplicity) to date statistical models (e.g. graphical models, log-linear models) do not usually distinguish the case $Y \perp\!\!\!\perp X|(Z = \mathtt{y})$ but $Y \not\perp\!\!\!\perp X|(Z = \mathtt{n})$ from the case $Y \not\perp\!\!\!\perp X|Z$. Interactive statistical graphics do sometimes explore the former through 'slicing'.

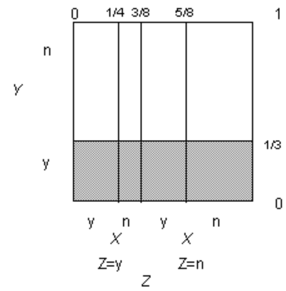

Figure 24: Random variables Y and X are conditionally independent given Z and Y and Z are conditionally independent given X. Unconditionally Y and X are independent, as are Y and Z. However, X and Z are dependent.

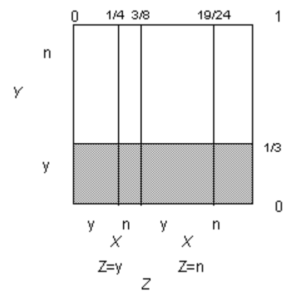

Figure 25: Random variables X, Y and Z are mutually independent.

While more could be said about conditional independence via eikosograms the essential points are made with the few we have already presented. Further exploration is beyond the scope of the present paper.

5 Diagrams for Probability Modelling.

Like probability, eikosograms presume that events or random variables have already been provided. Eikosograms are useful to explore the properties of particular probability models but are of no use in identifying the random variables or events on which the probabilities are defined. This aspect of probability modelling must be served by different diagrams.

One might think that this would be the proper place to use Venn diagrams, to define the events on which probability operates. However, Venn diagrams are ideally suited to describe *logical relationships between existing events*; what is needed are diagrams which help *define events* in the first place.

As is often the case, turning to historical sources where concepts were first correctly formulated can provide insight into how best to teach those concepts. After all, those earlier struggles are akin to those of students and, like students, those first formulating the concepts look for aids, diagrammatic and otherwise, which help naturally to clarify the concept itself.

5.1 Outcome trees.

Trees are perhaps the earliest diagrams used in probability dating back to at least Christiaan Huygen's use in 1676 (see Shafer, 1996). They are natural when the outcomes lead one to another in time. Figure 26(a) shows a simple tree describing two tosses of a coin. Branches at a point in the tree represent the mutually exclusive and exhaustive outcomes which could follow from that point.

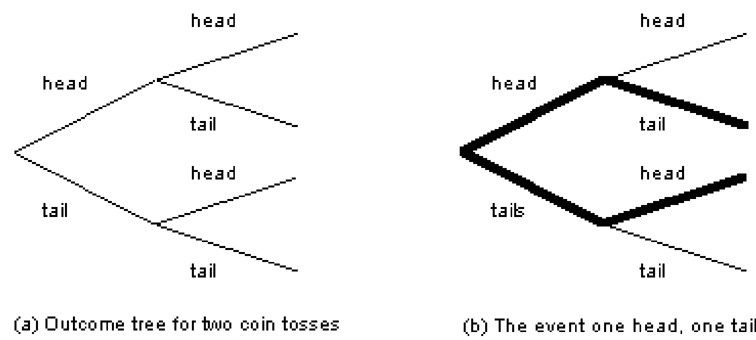

Figure 26: Defining events on an outcome tree.

While some notion of time is generally associated with movement from left to right across the tree, this is not strictly required. For some situations, the ordering of the tree branches might rather be one of convenience. For example, the tree of Figure 26 could also be used to provide a description for the simultaneous toss of two coins, with left and right components being labelled as "Coin 1" and "Coin 2".

Either way, the diagram provides a complete description of the situation under consideration in terms of all possible outcomes at each step – hence the name *outcome tree*.[5] If the branching probabilities were attached we would have the familiar *probability tree*. However, determining the probabilities is a separate stage in the probability modelling, and so it is best to spend some time with the outcome tree before moving on to this next stage.[6]

Events can now be defined by reference to the outcome tree. For example, the thick branches of Figure 26(b) show the event 'one head and one tail' without specifying which toss produced which. Similarly, if we were considering the event 'a head followed by a tail' only the topmost of the two thickly shaded paths would define the event; the bottommost of the two defines the event 'a tail followed by a head'. These two events combine to produce the first event of 'one head, one tail'.[7] The notion of *outcome space* (or more traditionally the *sample space*, a term we find to be less clear) could now be introduced as the set of all individual paths through the tree. An event, being a collection of paths, is simply a subset of the outcome space.

Outcome trees describe what can happen, step by step. The probability model is built on this structure by attaching conditional probabilities to each branch. The resulting probability tree will visually emphasize the conditional branching structure of the probability model whereas the corresponding eikosogram will visually emphasize the probability structure itself. One is easily constructed from the other since they contain the same information. The important difference is the different spatial priority each gives to the components of that information.

5.2 Outcome diagrams.

While outcome trees are often the most natural way to show *how* outcomes are possible, in some problems it is simpler just to show *what* outcomes are possible.

[5] Other authors, notably Edwards(1983) and following him Shafer (1996), prefer the name *event tree* for this diagram.

[6] Huygens's (1676) tree was not a probability tree in the modern sense. Huygens was interested in solving an early version of the gambler's ruin problem and labelled his branches with the 'hope' of winning (essentially the odds of winning at each stage) and the return due the gambler if the game were ended at that point. According to Shafer (1996, p.4) "[i]t was only after Jacob Bernoulli introduced the idea of mathematical probability in *Ars Conjectandi* that Huygens's methods became methods for finding 'the probability of winning'." (*Ars Conjectandi* was published posthumously in 1713.)

There are many interconnections between the players in this story. Jacob was the brother, teacher, and ultimately the mathematical rival of the Johann Bernoulli under whom Euler studied. Euler's father had attended Jacob's lectures and had lived with Johann at Jacob's house.

[7] This is the usual probabilistic use of the word event. Recently, in the development of a general theory for causal conjecture (one that depends heavily on the outcome tree description), Shafer has proposed calling such events *Moivrean events*. This then permits him to introduce what he calls *Humean events* to capture what common usage might consider to be a causal event in the tree structure. For example, the taking of a given branch might be considered the 'event' which 'caused' all that followed to be possible. The branch would be a Humean event whereas a Moivrean event must be one or more complete paths through the tree. With the introduction of Humean events for each branch, one can see why Shafer (1996) would choose to call these diagrams 'event trees'.

Since probability theory depends only on so-called 'Moivrean' events, we prefer 'outcome trees' to 'event trees'.

A notable early example of this approach is De Moivre's 1718 *Doctrine of Chances* in which he developed probability theory by addressing one problem after another. Although postdating Huygens (1676), no probability trees appear there. De Moivre did, however, find it convenient to completely enumerate all possible outcomes for some problems and, occasionally, to arrange these spatially in a table (e.g. De Moivre, 1756, p. 185). To each outcome, the number of 'chances' or frequency with which it can occur was attached and provided the information needed to determine the probability of any event composed from the listed outcomes.

In more modern times (dating to at least Fraser (1958) and predating standard use of Venn diagrams in probability books), it has been useful for teaching purposes to show all possible outcomes as spatially distinct points in a rectangular field as in Figure 27 (a). The spatial locations are arbitrary and so may be chosen so the events of interest

Figure 27: Defining events on an outcome diagram.

easily display as regions encompassing those outcomes which make up the event. In Figure 27(b) there are three non-overlapping regions which cover the entire field illustrating three mutually exclusive and exhaustive events. In Figure 27(c) two overlapping regions are drawn indicating two different events which have some outcomes in common.[8] In this figure, the unenclosed outcomes seem to constitute an event of no intrinsic interest; if they were of interest they would be best enclosed in a separate third region.

As with outcome trees, probabilities are missing from the outcome diagram. It is necessary to add them (usually to each individual outcome) in order to complete the probability model. Once outcome probabilities and events are in hand, any eikosogram for the events can be determined, although with more work than from a probability tree. Note however that, unlike probability trees, it will not generally be possible to construct an outcome diagram (and possibilities) from an eikosogram; at best only the construction of a Venn diagram (and attendant probabilities) will be possible.

5.3 A proposed teaching order.

The diagrams now in hand need to be used in concert to maximize their effectiveness in teaching probability and probability modelling.

Probability itself should be first introduced as an abstract concept related to area via eikosograms and further explored in the order delivered in Section 4. The focus should be on the mathematical abstraction of probability as grounded in a diagram with a simple raindrop metaphor. This material should be well exercised as preparation for its application. Those of mathematical bent could be drawn through the symbolic formalism of probability axioms based on conditional probability as defined by the eikosograms.

Outcome trees should then be introduced to provide the structure of a probability model for a real probabilistic situation. The real situation motivates the reasoned definition of a tree. This tree thus provides a situational description which can be used to define events and variables and so doing gives the student the first steps in understanding the probabilistic situation.

Next would be to assign branch probabilities which further model the situation. Given the probability tree, the corresponding eikosogram can be constructed and the probabilistic consequences of the model examined. Outcome trees and eikosograms would then be worked hand in glove to exercise much of probability theory in a variety of natural contexts. The challenge would be to come up with a variety of realistic problem situations to work on; this is easier done than coming up with realistic *probability* situations which sensibly exercise a Venn diagram.

[8] Figure 27(c) is also a diagram which would be useful to ground Venn's diagrams in an application and is often used for that purpose. It is a mistake, however common, to call Figure 27(c) a Venn diagram.

Outcome diagrams would be introduced last. In their discussion it should be pointed out that outcome diagrams are not generally as useful as outcome trees wherever the latter are applicable. For example, in the toss of two coins, the outcome diagram might have four outcomes – 'HH', 'TT', 'HT', and 'TH' – or it might only have three outcomes – '2H', '2T', and '1H,1T'. Only the first of these outcome diagrams would match the four paths of the outcome tree of Figure 26. Whenever an outcome tree is possible, it is recommended to be constructed first; the outcome space from the outcome tree (i.e. all of the paths through the tree) can be used to define the outcome diagram. Determining model probabilities for each of the points in an outcome diagram is often more difficult than determining the branching probabilities for an outcome tree.[9]

Events defined from an outcome diagram (perhaps constructed via an outcome tree) would then be used to explore the probability of one or another event occurring, of both events occurring, etc. as the situations warranted. In discussion of the logic of the intersection and union of events, only the outcome diagrams are needed. Venn diagrams (e.g. as in Figure 3) would be used only to introduce a further level of abstraction so as to discuss the logic more generally if that were desired.

6 Concluding remarks

Diagrams are important in learning any material, provided the diagram is well matched to that material. The eikosogram is just such a diagram for the introduction, definition, and exploration of probability and its attendant concepts such as conditional, marginal, and joint distributions as well as the more subtle concepts of probabilistic dependence and independence both unconditionally and conditionally.

Eikosograms obey Venn's dictum to match features of the diagram directly to the symbolic expression of the ideas. They fulfill Wittgenstein's notion of an 'ostensive definition' in that they can be used directly to define what is meant by these probability concepts. What eikosograms do not do is say how to use probability to model the real world.

This focus entirely on the mathematical abstraction of probability is a strength. Eikosograms permit a fundamental understanding of probability concepts to be had unclouded by the inherent difficulty of probability modelling. They do so by providing a definitive diagrammatic grounding for the symbolic expressions rather than one which appeals to some putatively natural application. Not only is the simultaneous introduction of probability and its application (often a source of confusion to many students) easily avoided but the important distinction between probability and model can be made early and more easily maintained thereafter.

If Venn's diagrams are to play a role in teaching probability it must be one considerably diminished from their present role. Outcome trees and probability trees have greater value for understanding events and the structure of a probability model. Eikosograms are coincident with probability. And outcome diagrams do much of the rest. Because of their inherent weaknesses for teaching probablity, it might be best at this time to avoid Venn diagrams altogether.

It is true that the intersecting ring diagrams are not original to Venn. But neither are they to Euler. The history of the diagrams, particularly in Christian symbolism, has shown them to be long associated with the demonstration of things separate and common to one another. This association is ostensibly inseparable from the diagrams. Given the religious training of both Euler and Venn, as well as the time periods in which these men lived, it seems likely that both men would have been aware of the vesica piscis and of the Christian symbolism associated with the two and three ring diagrams.

Euler's innovation was to use two-ring diagrams to demonstrate Aristotle's four fundamental propositions and to use more rings to illustrate the *known* outcomes of the syllogisms of Aristotelian logic. Venn, well aware of Euler's use, took the idea of intersecting rings (and of intersecting ellipses) to build a diagram which could be used to derive the consequence of possibly complex syllogisms as the logical information became available.[10] Each was an important and innovative use in its own right.

Historically and conceptually, eikosograms are direct descendants from Venn diagrams (e.g. Edwards, 1972). Their information content is that of probability and is easily organized and conveyed. Eikosograms should play a

[9] The example just given is a case in point. Early in the history of probability where it was applied to games of chance, Laplace's 'Principle of Indifference' was often applied to situations to model their probability. This principle says to model distinguishable outcomes as equiprobable. In the example just given, this would mean assigning equal probability of 1/4 to each of four outcomes in the first case and probabilities of 1/3 to each of three outcomes in the second. The latter solution was disposed of by applying the principle to the outcome tree thus assigning conditional probability of 1/2 to each of the two branches along the tree and so probability of 1/4 to each path in the tree.

[10] Venn even describes how to construct a physical apparatus based on the four ellipse diagram which can be used to carry out the logical calculations – foreshadowing today's digital, but electronic, computer.

central role in teaching probability. Venn diagrams can be safely set aside, their value replaced by outcome trees and outcome diagrams.

Appendix: Use of Venn diagrams in probability texts

Judging by today's texts, one might have thought that Venn diagrams had been used in expositions of probability for well over 100 years since Venn first wrote about them, or at least dating back to the beginnings of the use of an axiomatic set theoretic approach to probability. But as the following table shows, this doesn't seem to be the case.

Author	Date	Title	Use of Venn Diagrams
LaPlace	1812	Theorie Analytique des Probabilites	None
Venn, J.	1876	Logic of Chance	None
Venn, J.	1881	Symbolic Logic	Introduction and extensive use
Woodward, R.S.	1906	Probability and the Theory of Errors	None
Poincare, H.	1912	Calcul des Probabilites	None
Burnside, W.	1928	Theory of Probability	None
Jeffreys, H.	1939 1960	Theory of Probability (1st Ed.) Theory of Probability (3rd Ed.)	None None
Feller, W.	1950	An introduction to probability theory and its application (1st Ed.)	Yes
Kolmogorov, A.N.	1951	Foundations of the Theory of Probability (2nd Eng. Ed.)	None
Levy, P.	1954	Theorie de L'Addition des Variables Aleatoire	None
Loeve, M.M.	1955	Probability Theory: Foundations, random sequences	None
Cramer, H.	1955	The Elements of Probability Theory and Some of its Applications	None
Renyi, A.	1957	Calcul des Probabilites	None, but uses his own series of concentric circle diagrams to illustrate sets, their intersection and union
Fraser, D.A.S.	1957	Nonparametric Methods in Statistics	None
Fraser, D.A.S.	1958	Statistics: An Introduction	No. Instead he uses what we call *outcome diagrams* though he doesn't name them.
Dugue, D.	1958	Ensembles Mesurables et Probabilisables	None, but shows a (noncircular) set B nested within a larger (noncircular) set A
Derman, C.	1959	Prob. and Stat. Inference for Engineers	None, even though it begins with a set theoretic approach
Gnedenko, B.V.	1966 1968	Theory of Probability (3rd Ed.) Theory of Probability (4th Ed.)	Yes, but doesn't name them Yes, but now introduced with quotes as "so-called Venn diagrams"
David, F.N. and D.E. Barton	1962	Combinatorial Chance	None
Lindley, D.	1969	Intro. to Prob. and Stats. from a Bayesian Viewpoint	Not really, uses overlapping rectangular boxes for motivating axioms but curiously not for his conditional probability axiom

The table summarizes the presence or absence of Venn diagrams for several books. Many authors used no diagrams or used their own diagrams. Some, like Gnedenko (a student of Kolmogorov) used Venn diagrams without calling them such. In any case use of the diagrams in probability seems to have been rare and certainly not popular until more than 100 years after Venn promoted them for symbolic logic.

References

Baron, M.E. (1969). "A Note on the Historical Development of Logic Diagrams: Leibniz, Euler and Venn", *The Mathematical Gazette Vol. LIII*, pp. 113 -125.

Burkhardt, T. (1967). *Sacred Art in East and West*, Perennial, London.

Cleveland, W.S. (1985). *The Elements of Graphing Data*, Wadsworth, Monterey California.

DeMoivre, A. (1756). *The Doctrine of Chances, or, A Method of Calculating Probabilities of Events in Play: 3rd and Final Edition*, Reprinted 1967. Chelsea, New York.

Dewdney, A.K. (1999). *A Mathematical Mystery Tour*, John Wiley and Sons, New York.

Dunham, W. (1994). *The Mathematical Universe*, John Wiley and Sons, New York.

Edwards, A.W.F. (1972). *Likelihood*. Cambridge University Press, Cambridge.

Edwards, A.W.F. (1983). "Pascal's problem:the 'gambler's ruin'." *International Statistical Review*, 51, pp. 73-79.

Euler, L. (1761). Letters CII through CVIII of *Lettres à une Princesse d'Allemagne* dated Feb. 14 through March 7, 1761 appearing in English translation by D. Brewster as *Letters of Euler on different subjects in Natural Philosophy addressed to a German Princess, Vol. I* 1840, Harper & Brothers, N.Y. pp. 337 -366.

Feller, W. (1950). *An introduction to probability theory and its application* (1st Ed.), Wiley, NY.

Fraser, D.A.S. (1958). *Statistics: An Introduction*, Wiley, NY.

Friendly, M. (1994). "Mosaic displays for Multi-way contingency tables" *Journal of the American Statistical Association*, 89, pp. 190-200.

Friendly, M. (2002). "A brief history of the mosaic display." *Journal of Computational and Graphical Statistics*, 11. pp. 89-107.

Gardner, M. (1958). *Logic Machines and Diagrams*. McGraw-Hill, N.Y.

Glock, H-J. (1996). *A Wittgenstein Dictionary*. Blackwell, Oxford U.K.

Halley, E. (1693). "An estimate of the degrees of mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw, with an attempt to ascertain the price of annuities on lives." *Philosophical Transactions*, 17, pp. 596-610.

Heath, T.L. (1908). *The Thirteen Books of Euclid's Elements, Volume I (Introduction and Books, I, II)*, Cambridge University Press, Cambridge, U.K.

Huygens, C. (1676). "My last question on those matters that I published on Reckoning in Games of Chance; once proposed by Pascal" in English translation by G. Shafer in Appendix A. of *The Art of Causal Conjecture.*

Liungman, C.G. (1991). *Dictionary of Symbols*. ABC-CLIO, Santa Barbara, CA (also on-line at www.symbols.com)

Mann, A.T. (1993). *Sacred Architecture*, Element Books, Dorset UK.

O'Connor, J.J. and E.F. Robertson (2001). "The MacTutor History of Mathematics Archive", http://www-history.mcs.st-andrews.ac.uk/history/, School of Mathematics and Statstics, University of St. Andrews, Scotland.

Oldford, R.W. (1995). "A physical device for demonstrating confounding, blocking, and the role of randomization in uncovering a causal relationship". *The American Statistician*, 49 (2), pp. 210-216.

Oldford, R.W. (2001a). "Theorem of Pythagoras", http://www.math.uwaterloo.ca/navigation/ideas/grains/pythagoras.shtml.

Oldford, R.W. (2001b). "Constant from Randomness (estimating pi)", http://www.math.uwaterloo.ca/navigation/ideas/grains/pi.shtml.

Pennick, N. (1980). *Sacred geometry: symbolism and purpose in religious structures*, Turnstone Press, Wellingborough

Plato, "The Meno Dialogue" (translated by W.K.C. Guthrie), pp. 353-384 of *Plato: The Collected Dialogues (including the Letters)* edited by E. Hamilton and H. Cairns, Princeton University Press, Princeton, NJ.

Santalo, L.A. (1976). "Integral Geometry and Geometric probability", *Encyclopedia of Mathematics and its Applications*, Vol. 1, Addison-Wesley, Reading MA.

Shafer, G. (1996). *The Art of Causal Conjecture*. MIT Press, Cambridge MA.

Venn, J. (1866). *The Logic of Chance*, (2nd Edition, 1876) Macmillan and Co., London.

Venn, J. (1880). "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. X (Fifth Series) July, pp. 1- 18.

Venn, J. (1881). *Symbolic Logic*, (2nd Edition, 1894) as reprinted 1971 by Chelsea Publishing Co., New York.

Wittgenstein, L. (1964). *Philosophical Remarks*, University of Chicago Press, Chicago, IL.