Understanding probabilistic independence and its modelling via Eikosograms and graphs.

R.W. Oldford
University of Waterloo

January 17, 2003

Abstract

All possible independence structures available between three variables are explored via a simple visual display called an eikosogram (see Cherry and Oldford, 2002). Formal mathematical development is complementary rather than necessary.

If well understood, independence structures provide a solid basis for discussion of study design issues and statistical modelling. Eikosograms provide a simple visual basis for that understanding. Graphical and log-linear models, as well as covariance graphs are examined via eikosograms with a critical eye towards how these models either succeed or fail to capture the different independence structures.

New graphs are derived from the eikosograms which permit simple visual reasoning about the independence structures possible between three variables. Entirely visual proofs are available and given. Some independence structures which hold in general, are shown not to hold when one or more variables is binary. In this case some further results can be proven visually. Consequences for statistical planning of this and other results are briefly discussed.

1 Introduction

Cherry and Oldford (2002) show how the fundamentals of probability, its meaning, axioms, Bayes' and other theorems, conditional and unconditional independence, dependence relations, and the distinction between events and variables all naturally fall out of a particular diagram called an eikosogram. In this paper eikosograms are used to explore more deeply the related notions of conditional and unconditional independence. In particular, the variety of independence structures available with three categorical variables is explored in detail.

Much of the motivation is to provide simple visual helps which will enable one to better appreciate and to understand the complexity of dependence and independence relationships which are possible between random variables. This topic is important not only for understanding probability but also for understanding statistical modelling and design. Eikosograms, and the new graphs derived from them, make it possible to introduce these ideas and to prove results about the relations between independence structures without resort to symbolic mathematics.

Such an understanding can then be exploited in discussion of statistical design and modelling. The eikosograms provide a visual display of the underlying probability distribution against which the characteristics of statistical models can be fruitfully compared. Critical comparison reveals the strengths and weaknesses of the model and leads to a deeper understanding of the models and their application. Many planning and design fundamentals can be expressed in terms of the relation between three variates – a response variate, a treatment variate, and some auxiliary (i.e. non-treatment) explanatory variate which may or may not be known. Understanding the independence structures for three variables gives insight into such fundamentals as blocking or matching and randomization. A new result proved here shows here that extra care needs to be taken when the auxiliary explanatory variate is binary.

The remainder of the paper is organized as follows. Secton 2 gives a brief introduction to eikosograms and how they display the dependence between variables and between some variables conditional on others. The absence of dependence (unconditionally or conditionally) visually stands out in an eikosogram and visually defines what is meant by independence. Dawid's (1979) notation is extended slightly from a binary to n-ary operator to describe complete, or mutual, independence between all of its arguments. This helps clarify the variety of independence constraints which must hold to produce complete independence.

Section 3 enumerates all independence structures (there are eight modulo permutation of the variables, seven if all are binary) which can exist between three categorical variables. These independence structures are easily read off the eikosograms (numerical details of which are recorded in the Appendix) and the eikosograms can be used to illustrate and to prove theorems about independence relationships.

Section 4 deals with the common statistical models for categorical data and how they match up (or not) to the variety of independence structures uncovered in Section 3. Section 4.1 explores these structures via graph-based models – both those based only on marginal independence relations (e.g. covariance graphs of Cox and Wermuth, 1996) and those based on conditional independencies, the so-called graphical models (e.g. Darroch et al 1980). When only three variables are considered, combining both types of graphs in a single one yields a one to one correspondence between the graphs and the possible independence structures displayed in Section 3. Being a proper super-set of the graphical models, the hierarchical log-linear models of Section 4.2 have many of the same shortcomings but also some advantages over graphical models in describing independence and dependence structures of statistical interest.

Section 5 adapts the combined graphs from Section 4 to show how theorems on independence between three variables can be easily proven using a graph where known independence or dependence is explicitly marked with distinct edges – absence of edges asserts only ignorance. Similar in spirit to Venn's use of his diagrams in logic (e.g. see Cherry and Oldford 2002), these graphs can be used to visually express and even to prove independence relations without resort to formal mathematics; Dawid's notation provides the symbolic expressions. Eikosograms are used to give visual proof of two basic results on independence from which the remaining results can be proved using only the graphs. The case where at least one of the three variables is binary requires formal proof as detailed in the Appendix. Altogether, the graphs can be used to prove that the structures seen in Section 3 are all that are possible. Moreover, how the graphs help shed light onto important and fundamental statistical design issues is briefly discussed.

The last section contains a few concluding remarks.

2 Eikosograms

Eikosograms display probability for discrete random variables as area on a unit square; the calculation rules for probability are one to one with the calculation rules for rectangular areas (see Cherry and Oldford, 2002). Figure 1 shows three eikosograms displaying the marginal distribution of a binary random variable Y and its joint probability

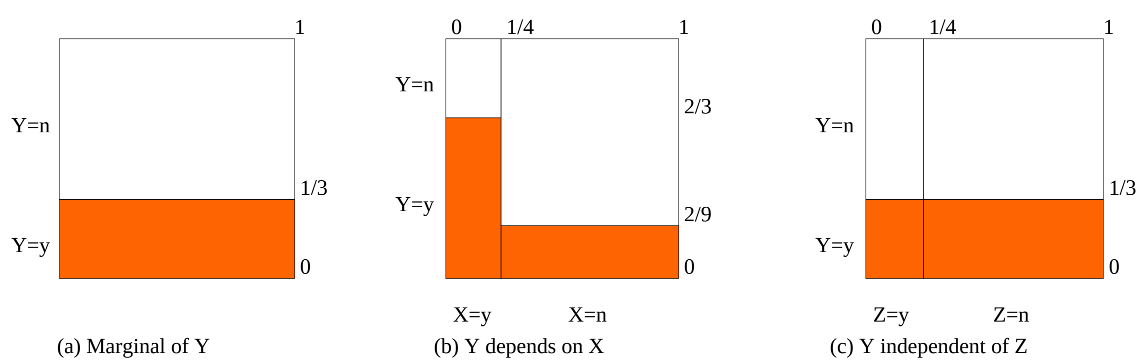


Figure 1: The eikosogram for one and for two variables.

distribution with each of two other binary random variables, X and Z respectively (the values '$\mathtt{y}$' and '$\mathtt{n}$' stand for the binary possibilities '$\mathtt{yes}$' and '$\mathtt{no}$'). The area of the shaded areas in all three diagrams is $1/3$ and so $Pr(Y = \mathtt{y}) = 1/3$ in each of the three cases.

In Figure 1(b), one can additionally read off from the horizontal axis that $Pr(X = \mathtt{y}) = 1/4$ (the area of the entire vertical strip) and that $Pr(X = \mathtt{n}) = 1 - 1/4 = 3/4$; from the vertical axis we read the values of the conditional probabilities $Pr(Y = \mathtt{y}|X = \mathtt{y}) = 2/3$ and $Pr(Y = \mathtt{y}|X = \mathtt{n}) = 2/9$. Because these conditional probabilities are not equal, there is probabilistic dependence between the random variables Y and X.

To find the marginal of Y as in Figure 1(a) from the joint of Y versus X as in Figure 1(b) a visual metaphor helps. Imagine Figure 1(b) as a container of water divided into two chambers by a barrier at $1/4$ with the shaded regions representing the water in each chamber. Finding the marginal distribution amounts to removing the barrier and allowing the water to settle at its natural level – here the depth of $1/3$ across the container as in Figure 1(a).

The relationship between Y and Z is different. In Figure 1(c), we see that $Pr(Y = \mathsf{y}|Z = \mathsf{y}) = Pr(Y = \mathsf{y}|Z = \mathsf{n}) = 1/3 = Pr(Y = \mathsf{y})$, and similarly for $Y = \mathsf{n}$, which means that Y and Z are probabilistically independent. Visually, this independence is indicated by the flatness along the shaded regions in the eikosogram of Y and Z. In Figure 1(c) removal of the barrier at $1/4$ will not affect the water level – hence independence.

Figure 2 shows different levels of association between the two binary variables Y and X. Across these diagrams

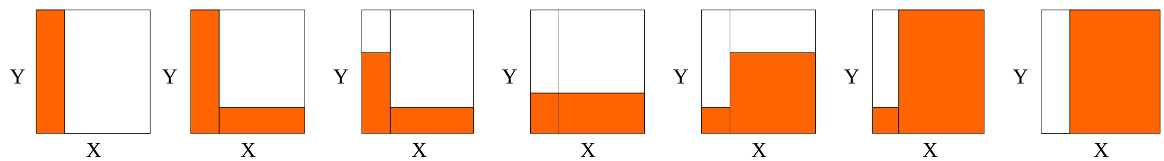
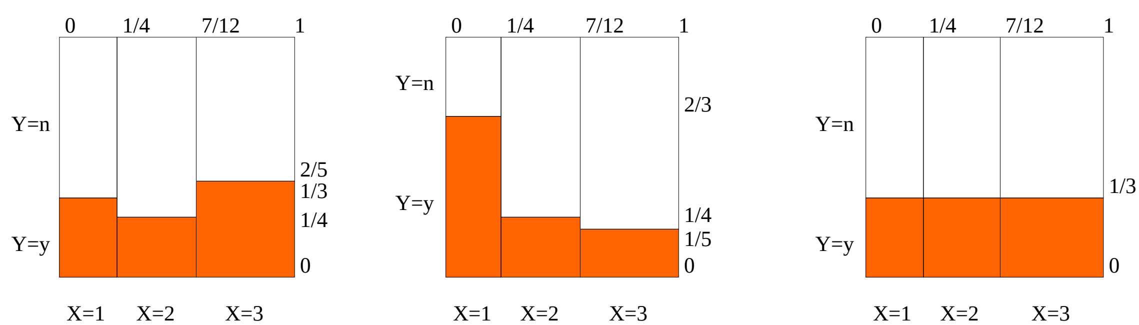


Figure 2: Binary associations left to right. (a) Perfect positive association: $Y = X$; (b) Part perfect positive association: $1 = Pr(Y = \mathsf{y}|X = \mathsf{y}) > Pr(Y = \mathsf{y}|X = \mathsf{n})$; (c) Positive association: $Pr(Y = \mathsf{y}|X = \mathsf{y}) > Pr(Y = \mathsf{y}|X = \mathsf{n})$; (d) Independence; (e) Negative association: $Pr(Y = \mathsf{y}|X = \mathsf{y}) < Pr(Y = \mathsf{y}|X = \mathsf{n})$; (f) Part perfect negative association; (g) Perfect negative association: Y is the complement of X.

the marginal distribution of X is held fixed, that of Y is not. Providing a measure of association is not obvious; even the most commonly recommended one, the odds ratio, fails to distinguish 'Perfect' association (Figure 2a) from 'Part Perfect' association (Figure 2b) (or "Absolute" from "Complete" association as in Fienberg, 1977, pp. 18-19).

The eikosogram naturally distinguishes the variable on the vertical axis making the reading of its conditional probabilities easy. Two eikosograms, one with Y on the vertical axis and one with X on the vertical axis would be required to present the variables symmetrically. Probabilistic independence will produce a flat shaded region in both eikosograms. For binary random variables X and Y the associations seen in Figure 2 will be apparent in both eikosograms for $Pr(Y = \mathsf{y}|X = \mathsf{y}) > Pr(Y = \mathsf{y}|X = \mathsf{n})$ holds iff $Pr(X = \mathsf{y}|Y = \mathsf{y}) > Pr(X = \mathsf{y}|Y = \mathsf{n})$ – similarly for negative association.

Variables displayed in eikosograms need not be binary. Figure 3 shows three eikosograms with a vertical binary



(a) Y depends on X (b) Pr(Y=y|X) declines with X (c) Y independent of X

Figure 3: Multiple categories for the conditioning variable.

variable, Y, and a three-valued ordinal variable X on the horizontal axis. Probabilities can be read off as before and flatness again indicates probabilistic independence while absence of flatness indicates dependency between the two variables. If the vertical variable has multiple categories distinguished by different colours, independence would correspond to horizontal colour bars across the eikosogram.

When more than two variables are involved, all but one are displayed along the horizontal axis as conditioning variables. Figure 4 shows a typical eikosogram for three binary variables together with the canonical ordering we will

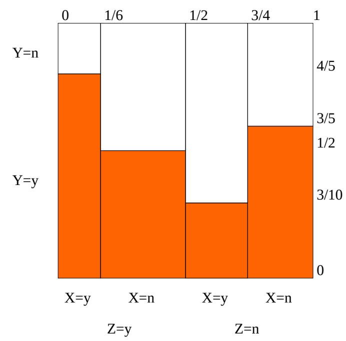


Figure 4: Three variables (binary here); Two conditioning variables. Y vs. X & Z.

adopt for the values of the binary conditioning variables. As with two variables, the conditional probabilities can be read off the vertical heights of the shaded bars to give:

$$\begin{aligned} Pr(Y=\mathsf{y}|X=\mathsf{y},Z=\mathsf{y}) &= 4/5, \quad Pr(Y=\mathsf{y}|X=\mathsf{n},Z=\mathsf{y})=1/2, \\ Pr(Y=\mathsf{y}|X=\mathsf{y},Z=\mathsf{n}) &= 3/10, \quad Pr(Y=\mathsf{y}|X=\mathsf{n},Z=\mathsf{n})=3/5\ . \end{aligned}$$

Similarly reading from the horizontal axis gives $Pr(X=\mathsf{y},Z=\mathsf{y})=1/6$, $Pr(Z=\mathsf{y})=1/2$, and so on.

Again dependence and independence relationships are apparent from the eikosogram but, with three and more variables, these can be much more involved. Figure 5 gives some indication of the possibilities. Here again there

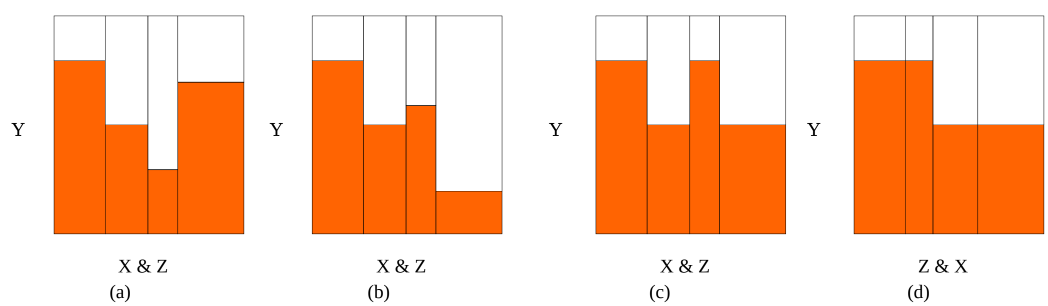


Figure 5: Various dependency relations (a) inconsistent dependence, (b) consistent dependence, (c) conditional independence, (d) as in (c) but $X\&Z$ have their order in the display interchanged.

are three binary variables and flatness is an indicator of some sort of independence. The values of the conditioning variables are canonically ordered with the values of the first variable mentioned changing most quickly (y to n) left to right across the eikosogram as in Figure 4.

In Figure 5(a) we say that the dependence is inconsistent because the conditional probability that $Y=\mathsf{y}$ given the values of the others decreases as the X value changes from y to n when $Z=\mathsf{y}$ and does the opposite when $Z=\mathsf{n}$. The inconsistency is extreme in the sense that it is also the case that the conditional probability of $Y=\mathsf{y}$ given the values of the others decreases as the Z value changes from y to n when $X=\mathsf{y}$ and does the opposite when $X=\mathsf{n}$. Either one or both of these inconsistencies could hold in any given situation.

In Figure 5(b) we say that the dependence is consistent. The conditional probability of $Y=\mathsf{y}$ given the values of the others changes with the value of X in the same direction (i.e. decreases) when $Z=\mathsf{y}$ and when $Z=\mathsf{n}$. A similarly consistent relationship holds between Y and Z when $X = \mathsf{y}$ and $X = \mathsf{n}$ (in Figure 5(b) the conditional probability decreases as Z changes from y to n when $X = \mathsf{y}$ and when $X = \mathsf{n}$). Either one or the other or both of these consistent dependencies could hold. Because both hold in Figure 5(b), the dependency is, in a sense, maximally consistent.

Figures 5(c) and 5(d) show a relationship between the three variables which is identical in both figures. The only difference is that in Figure 5(c) the values of X change first and in Figure 5(d) the columns of the eikosogram are re-ordered so that the values of Z now change first. The probabilistic relationship is identical, but the flats which appear in the eikosograms of both Figure 5(c) and Figure 5(d) are more obvious in the arrangement of the latter.

As before, these flats are indicative of some probabilistic independence. From either Figure 5(c) or 5(d), the two flat regions occur when $X = \mathsf{y}$ and when $X = \mathsf{n}$, each across all values of Z. The existence of these flat regions are necessary and sufficient conditions to conclude that Y and Z are conditionally independent given X, written as $Y \perp\!\!\!\perp Z|X$ (following Dawid, 1979).

Had it been the case in Figure 5(d) that a flat existed only on the left where $X = \mathsf{y}$ and not on the right where $X = \mathsf{n}$, then we would have Y conditionally independent of Z given the *event* $X = \mathsf{y}$ but conditionally *dependent* given the event $X = \mathsf{n}$. We write these two possibilities as $Y \perp\!\!\!\perp Z|\{X = \mathsf{y}\}$ and $Y \not\perp\!\!\!\perp X|\{X = \mathsf{n}\}$, respectively. The eikosogram allows us to visually distinguish the case for events from the corresponding case for variables (see Cherry and Oldford, 2002 for further discussion on this point).

From Figure 5 we can see that the following results hold:

$$
\begin{array}{lcl}
Y \perp\!\!\!\perp Z|X & \Longleftrightarrow & \text{both} \quad Y \perp\!\!\!\perp Z|\{X = \mathsf{y}\} \quad \text{and} \quad Y \perp\!\!\!\perp Z|\{X = \mathsf{n}\}, \\
Y \perp\!\!\!\perp Z|X & \not\Longrightarrow & Y \perp\!\!\!\perp Z.
\end{array}
$$

A formal proof can be easily had by symbolically describing the appropriate features of the eikosogram.

Figure 6(a) shows a distribution where the flat extends across all values of the conditioning variables. As was the

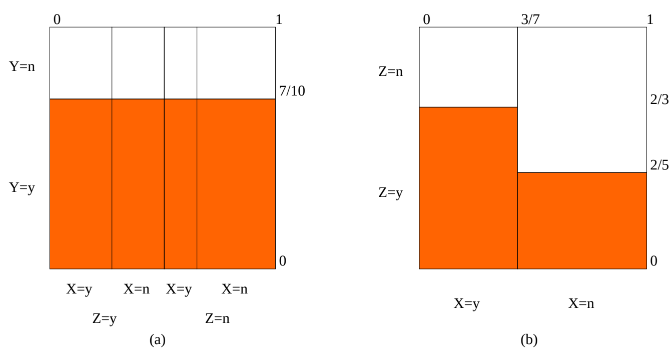


Figure 6: One 4-flat does not imply mutual independence of Y, X, and Z.

case with two variables, one might be tempted to think this complete flat meant the mutual independence of all three variables Y X and Z – but this would be wrong! From Figure 6(a) the eikosogram of the joint distribution of Z versus X can be derived and as shown in Figure 6(b) clearly demonstrates the dependence between Z and X.

A flat extending across all values of the conditioning variables as in Figure 6(a) implies the following four independencies: $Y \perp\!\!\!\perp X|Z$, $Y \perp\!\!\!\perp Z|X$, $Y \perp\!\!\!\perp X$, and $Y \perp\!\!\!\perp Z$. Moreover, the eikosograms of Figures 6(a) and (b) give visual proof of the following results:

$$
\begin{array}{lcl}
Y \perp\!\!\!\perp X|Z \quad \text{and} \quad Y \perp\!\!\!\perp Z|X \quad \text{together} & \Longrightarrow & Y \perp\!\!\!\perp X \quad \text{and} \quad Y \perp\!\!\!\perp Z \quad \text{but} \\
Y \perp\!\!\!\perp X|Z \quad \text{and} \quad Y \perp\!\!\!\perp Z|X \quad \text{together} & \not\Longrightarrow & X \perp\!\!\!\perp Z.
\end{array}
$$

It should come as no surprise that if, in addition to the premises above, we also have $X \perp\!\!\!\perp Z$, then it will follow that Y, X, and Z are mutually, or completely, independent.

2.1 A notation for complete independence

Expressing this last result in terms of Dawid's (1979) notation is difficult because the symbol $\perp\!\!\!\perp$ as used there is a binary operator, although on possibly vector valued operands. For example, the expression $X \perp\!\!\!\perp Y \perp\!\!\!\perp Z$, which might mistakenly suggest the mutual independence of all three variables, can be meaningfully interpreted only as $X \perp\!\!\!\perp Y$ and $Y \perp\!\!\!\perp Z$. A suggestive and compact notation for mutual independence can be had by allowing $\perp\!\!\!\perp$ as an operator on two or more symmetrically treated variables whose meaning is taken to be as follows:

$$
\begin{array}{rll}
\perp\!\!\!\perp(Y,X) & \text{means} & Y \perp\!\!\!\perp X \\
\perp\!\!\!\perp(Y,X)|Z & \text{means} & Y \perp\!\!\!\perp X|Z \\
\perp\!\!\!\perp(Y,X,Z) & \text{means} & \perp\!\!\!\perp(Y,X)|Z,\ \perp\!\!\!\perp(Y,Z)|X,\ \perp\!\!\!\perp(X,Z)|Y,\ \perp\!\!\!\perp(Y,X),\ \perp\!\!\!\perp(Y,Z),\ \text{and}\ \perp\!\!\!\perp(X,Z)\ \text{all hold} \\
 & & \text{or in the original (infix) notation} \\
 & & Y \perp\!\!\!\perp X|Z,\ Y \perp\!\!\!\perp Z|X,\ X \perp\!\!\!\perp Z|Y,\ Y \perp\!\!\!\perp X,\ Y \perp\!\!\!\perp Z,\ \text{and}\ X \perp\!\!\!\perp Z\ \text{all hold}
\end{array}
$$

When only the binary operation is called for, either the prefix or infix notation will be used depending on which seems clearer in the context.

As always, conditioning variables or events appear to the right of a vertical line (if more than one variable is conditioned on, as in Dawid (1979), they can appear listed separately within parentheses after the vertical bar). The prefix notation $\perp\!\!\!\perp(Y, \cdots, Z)$ is intended to indicate the complete probabilistic independence of its arguments which, as with pairwise independence, can occur for its arguments either unconditionally or conditionally given other random variables or events. So,

$$
\begin{array}{rll}
\perp\!\!\!\perp(Y,X,Z)|W & \text{means} & \perp\!\!\!\perp(Y,X)|(Z,W),\ \perp\!\!\!\perp(Y,Z)|(X,W),\ \perp\!\!\!\perp(X,Z)|(Y,W), \\
 & & \perp\!\!\!\perp(Y,X)|W,\ \perp\!\!\!\perp(Y,Z)|W,\ \text{and}\ \perp\!\!\!\perp(X,Z)|W \\
 & & \text{all hold}
\end{array}
$$

For four variables the recursive definition is

$$
\begin{array}{rll}
\perp\!\!\!\perp(Y,X,Z,W) & \text{means that} & \perp\!\!\!\perp(Y,X,Z)|W,\ \perp\!\!\!\perp(Y,X,W)|Z,\ \perp\!\!\!\perp(Y,Z,W)|X,\ \perp\!\!\!\perp(X,Z,W)|Y, \\
 & & \perp\!\!\!\perp(Y,X,Z),\ \perp\!\!\!\perp(Y,X,W),\ \perp\!\!\!\perp(Y,Z,W),\ \text{and}\ \perp\!\!\!\perp(X,Z,W), \\
 & & \text{all hold.}
\end{array}
$$

The extension of the notation to more than four variables is entirely analogous. $\not\perp\!\!\!\perp(Y, \cdots, Z)$ means that at least one of the independencies on the right hand side does not hold.

Given the number of independencies that are entailed, the phrase *complete* independence seems more evocative of the strength of the assertion than does the traditional *mutual* independence.

We are now in a position to succinctly express the result on complete independence which was suggested (but not proved) by Figure 6, namely:

$$\perp\!\!\!\perp(Y,X,Z) \iff \perp\!\!\!\perp(Y,X)|Z,\ \perp\!\!\!\perp(Y,Z)|X,\ \text{and}\ \perp\!\!\!\perp(X,Z) \tag{1}$$

3 Exploring independence for three variables

A collection of three or more random variables can have a variety of different independencies of potential interest without ever achieving complete independence. But not all combinations of pairwise independencies (conditional and unconditional) are possible. In this section we enumerate this set of possibilities by looking at the different ways in which flat regions describe independence with three variables.

Every flat region corresponds to an independency of some sort; absence of a flat region means a dependence. Flat regions which cross all values of the conditioning variables either as a single flat or as a sequence of plateaus (one plateau for each value of one of the conditioning variables) correspond to some independence of variables either unconditionally or conditionally. From these features, it is possible to see at once what the independence relations are. Moreover, it will be possible to use the eikosograms to make apparent a number of theoretical results.

With three variables, there are six different possible three-way eikosograms: each of the three variables can appear on the vertical and for each of these the order of the remaining two variables on the horizontal axis can be interchanged. In general there will be $n!$ different n-way eikosograms for n variables. Because we are looking for flat areas however, if care is taken to recognize that the discontiguous flats of Figure 5(c) would actually appear as contiguous were the horizontal variables interchanged as in Figure 5(d), then it is sufficient to draw only three (more generally n) of the possible eikosograms.

For three variables, when interest lies in examining conditional independence of two variables given a third, only three possible configurations of a three way eikosogram are of interest: a "full-flat" which means a single flat across all vertical bars; a multi-flat where for each value of one conditioning variable a separate plateau crosses all values of the other; or the absence of either of these, a "no-flat". These possibilities have been shown for three binary variables: a 'full-flat' in Figure 6 (a), a multi-flat in Figures 5(c) or (d), and a 'no-flat' in Figures 5(a) or (b).

For simplicity of exposition, it will be convenient to have all three variables be binary. These possibilities are now more precisely described as a '4-flat' a '2×2 flat' and a 'no-flat', respectively. Unless otherwise indicated, the results which follow hold for any number of categories for each of the three variables.

For any triple of variables Y, X, and Z, only three eikosograms need be considered; the canonical arrangements will be: Y vs. $X\&Z$, Z vs. $X\&Y$, and X vs. $Z\&Y$. Treating the variables symmetrically, it turns out that only four collections of flat configurations are possible. These are now examined in turn.

3.1 Case 1: All three diagrams are flat.

This is the case of complete independence, $\perp\!\!\!\perp(Y, X, Z)$. An example is given in Figure 7 (numerical values for this

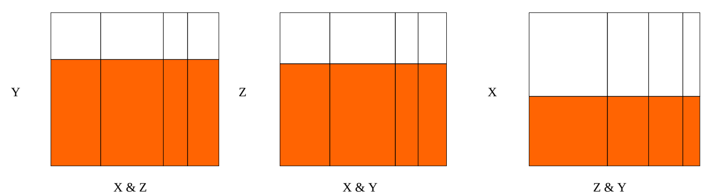


Figure 7: Complete independence: any two 4-flats imply the third is a 4-flat; three 4-flats if and only if Y X and Z are completely independent. Left to right: (a) Y vs $X\&Z$, (b) Z vs $X\&Y$, and (c) X vs $Z\&Y$.

and other examples are recorded in a table in the Appendix). The three variables are completely independent iff all three diagrams are flat. As indicated in the caption, if only two of the three eikosograms are produced and both show a 4-flat, then there is no need to produce the third, for it too must be a 4-flat.

Figure 7(a) shows $Y \perp\!\!\!\perp X|Z$ and $Y \perp\!\!\!\perp Z|X$ and, as a consequence, that $Y \perp\!\!\!\perp X$ and $Y \perp\!\!\!\perp Z$; similar results hold from Figures 7(b) and (c). That two 4-flats imply the third is a restatement of the result given by equation number (1); Figure 7(a) provides the first two results of the right hand side of (1), Figure 7(b) shows the third holds as well. It follows then that $\perp\!\!\!\perp(Y, X, Z)$ and so the third eikosogram must be flat as well.

In the case of complete independence, the pairwise independence relations are apparent (via the water container metaphor) from the three-way eikosograms. When complete independence does not hold, the marginal relationship between any pair of variables will need to be examined directly from the relevant two-way eikosogram.

3.2 Case 2: one 4-flat, two 2×2-flats

As seen in Figure 6 and in more detail here again in Figure 8, it is possible to observe a single 4-flat without there being complete independence between the variables. The two conditional independencies given by a 4-flat imply that the vertical variable (in Figure 8 this is X) is independent of every other variable both conditionally and unconditionally. The water container metaphor makes this plain and easily extends the result to any number of variables. In Figure 8 all conditional and marginal independencies are immediate consequences of the 4-flat.

For three variables, any independence between the remaining two variables, that is either $Y \perp\!\!\!\perp Z|X$ or $Y \perp\!\!\!\perp Z$, will imply complete independence between those three variables or $\perp\!\!\!\perp(Y, X, Z)$. (This is true for four or more variables

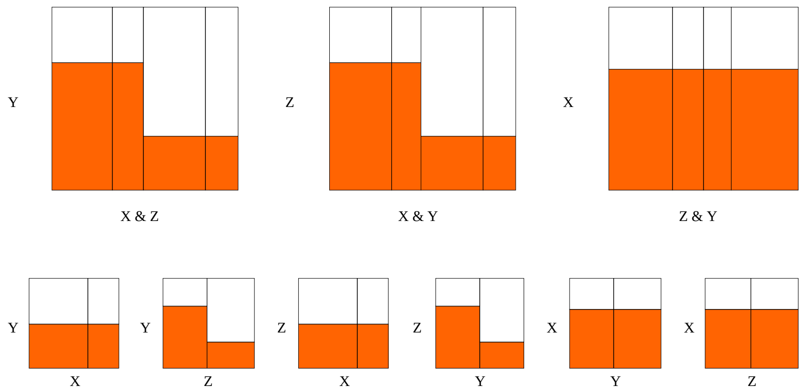


Figure 8: One 4-flat and two 2×2-flats.

provided the additional independence involves only the three named variables.) That is, both of the other three-way eikosograms must be 4-flats as well. The result is again easily seen from the eikosograms of Figure 8 via the water container metaphor. Symbolically,

$$X \perp\!\!\!\perp Z|Y, \quad X \perp\!\!\!\perp Y|Z, \quad \text{and either} \quad Y \perp\!\!\!\perp Z|X \quad \text{or} \quad Y \perp\!\!\!\perp Z \Longrightarrow \perp\!\!\!\perp(Y, X, Z)$$

It follows from the above that if a 4-flat is observed in one three-way eikosogram, then the others are either both 4-flats (Case 1), or both 2×2 flats (Case 2) and that it is impossible for all three-way eikosograms to be 2×2 flats.

It is possible, however, to have only two 2×2-flats with the third eikosogram containing no flats whatsoever; this is Case 3.

3.3 Case 3: two 2×2-flats, one no-flat

This happens when there is only one conditional-independence. Figure 9 shows an example in which the first and third

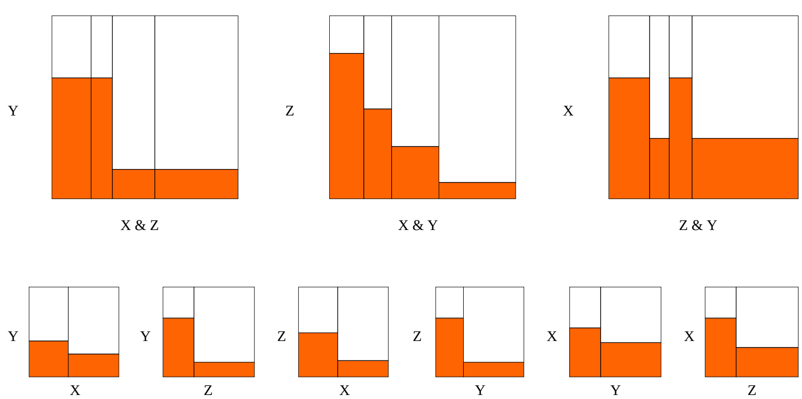


Figure 9: Two 2×2-flats, one no-flat.

eikosograms in the top row are each a 2×2 flat representing a single conditional independence, namely $Y \perp\!\!\!\perp X|Z$. No other independence, conditional or marginal appears to exist.

More formally,

$$Y \perp\!\!\!\perp X|Z \quad \text{and} \quad Y \not\perp\!\!\!\perp Z|X \Longrightarrow Y \not\perp\!\!\!\perp Z. \tag{2}$$

This result is easily proven via the eikosograms by recalling that the marginal eikosogram of Y versus Z is had from that of Y versus X and Z simply by removing the barriers between the Xs separately for each value of Z. As in Figure 9, when $Y \perp\!\!\!\perp X|Z$ there would be no effect on the corresponding water levels and so $Y \not\perp\!\!\!\perp Z$ if $Y \not\perp\!\!\!\perp Z|X$.

By symmetry (simply exchanging X and Y) or from the corresponding eikosogram it also follows that

$$Y \perp\!\!\!\perp X|Z \quad \text{and} \quad Z \not\perp\!\!\!\perp X|Y \Longrightarrow X \not\perp\!\!\!\perp Z$$

which explains the X vs Z marginal dependency observed.

The third dependency, namely $Y \not\perp\!\!\!\perp X$, does not always hold unless, as in this case, Z is binary. Then, when Z is binary (either or both of Y and Z can have more than two categories), the following must hold

$$Y \perp\!\!\!\perp X|Z \;\; , \;\; Y \not\perp\!\!\!\perp Z|X \;\; \text{and} \;\; Z \not\perp\!\!\!\perp X|Y \quad \Longrightarrow \quad Y \not\perp\!\!\!\perp X. \tag{3}$$

This explains the marginal dependence of Y and X seen in Figure 9. When (3) holds, a number of other results also follow. These are discussed later and the algebraic proof for the binary case is given in the Appendix.

That the result (3) does not hold when Z is not binary, is shown by the counter-example given by the three-way eikosogram of Figure 10 The numerical values are given in the Appendix.

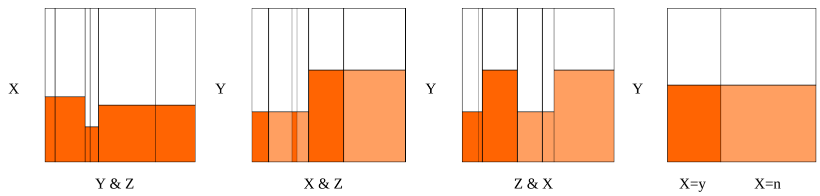


Figure 10: When Z is **not** binary $Y \perp\!\!\!\perp X|Z$, $Y \not\perp\!\!\!\perp Z|X$ and $Z \not\perp\!\!\!\perp X|Y \not\Longrightarrow Y \not\perp\!\!\!\perp X$. Left to right: (a) $X \not\perp\!\!\!\perp Z|Y$ and $X \perp\!\!\!\perp Y|Z$, (b) Again $Y \perp\!\!\!\perp X|Z$, $Y = \mathsf{y}$ darker when $X = \mathsf{y}$ than when $X = \mathsf{n}$, (c) Rearrange (b) to group first by $X = \mathsf{y}$ then $X = \mathsf{n}$, and finally (d) Remove the Z barriers in (c) to level out at $Y \perp\!\!\!\perp X$.

Observing one conditional independence between two variables, say $Y \perp\!\!\!\perp X|Z$, says nothing about any other conditional independence nor about any marginal independence. Any one of Cases 1 through 3 could have occured. However if no other conditional independence exists between the three variables, then it follows that the two conditionally independent variables Y and Z are each marginally dependent upon the conditioning variable Z. If, additionally, the conditioning variable is binary, then no marginal independence will hold. If Z is not binary, then Y and X might or might not be independent.

3.4 Case 4: three no-flats

It is possible for there to be no flat, neither any 4-flat nor any 2×2 flat, in any of the three-way eikosograms. Contrary to what might at first seem to be meant by this, it is not the case that this means that the three variables are completely dependent. Figure 11 provides an example where it is the case that no independence relationship exists between any variables, conditionally or unconditionally.

Like complete independence, the complete dependence illustrated in Figure 11 is everywhere between all the variables. For two variables, complete dependence is simply unconditional pairwise dependence and can be indicated symbolically by the symbol for the absence of independence: $\not\perp\!\!\!\perp$. For three or more variables, something different is needed as $\not\perp\!\!\!\perp(Y, \cdots, Z)$ does not preclude the possibilitiy of independence structure existing amongst the arguments, only that it is not complete.

We propose to use n-ary symmetric operator $\bowtie$ (as produced by the Latex symbol `\Join`) to indicate complete dependency; its negation $\not\bowtie$ will indicate the absence of complete dependence. The definition of $\bowtie(Y, \cdots, Z)$ follows the same pattern as the definition of $\perp\!\!\!\perp(Y, \cdots, Z)$ except that $\perp\!\!\!\perp$ is replaced everywhere by $\bowtie$. For the n-ary case with $n \geq 3$, it should be clear that

$$\begin{array}{lll} & \perp\!\!\!\perp(Y, \cdots, Z) \Longrightarrow \not\bowtie(Y, \cdots, Z), \quad \text{and} & \bowtie(Y, \cdots, Z) \Longrightarrow \not\perp\!\!\!\perp(Y, \cdots, Z) \\ \text{but} & \not\bowtie(Y, \cdots, Z) \not\Longrightarrow \perp\!\!\!\perp(Y, \cdots, Z) \quad \text{and} & \not\perp\!\!\!\perp(Y, \cdots, Z) \not\Longrightarrow \bowtie(Y, \cdots, Z). \end{array}$$

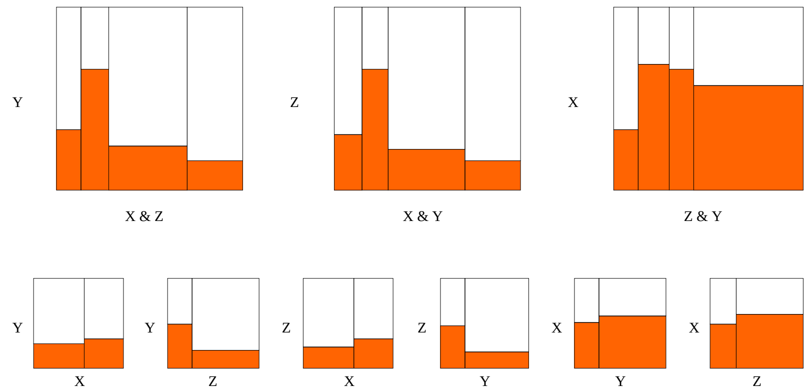


Figure 11: No flats. Complete dependence: $\bowtie(Y, X, Z)$

When $n = 2$, the implications in the second line above hold – $Y \perp\!\!\!\perp X$ is the same as $Y \not\bowtie X$ and $Y \not\perp\!\!\!\perp X$ the same as $Y \bowtie X$. This notation will be most useful when four and more variables are involved and different kinds of complete dependencies and complete independencies, conditional and unconditional, can be distinguished.

Figure 12 gives an example where there are no flats in the three-way eikosograms, indicating no conditionally

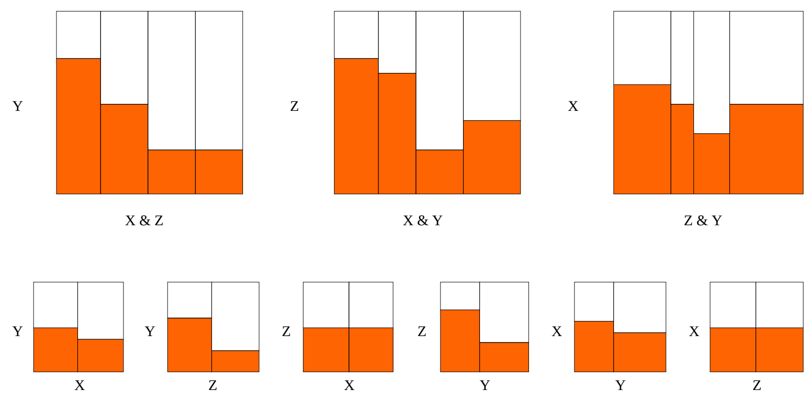


Figure 12: No flats. No complete dependence $\not\bowtie(Y, X, Z)$. One marginal independence. No other independence

independent distribution for any two variables given the third. Yet there is a single marginal dependence, namely $Z \perp\!\!\!\perp X$ as can be seen from the relevant flats in the second row of Figure 12. This example also has a conditional independence for one value of Z, namely $Y \perp\!\!\!\perp X | \{Z = \mathsf{n}\}$; this is not necessary to produce the observed marginal independence but rather included to show that further independencies exist and are of possible interest besides those associated with conditioning on every value of a random variable.

Figure 13 illustrates the case where no conditional independence exists but two marginal independencies do. This can appear surprising when encountered for the first time symbolically as in

$$Y \perp\!\!\!\perp Z \text{ and } Z \perp\!\!\!\perp X \not\Longrightarrow \perp\!\!\!\perp(Y, X, Z).$$

However, the possibility is clear from the eikosograms of Figure 13.

Figure 14 shows the case where all three marginal independencies occur but no other independence structure exists. To those just learning about probability this result expressed symbolically as

$$Y \perp\!\!\!\perp Z \text{ and } Z \perp\!\!\!\perp X \text{ and } Y \perp\!\!\!\perp X \not\Longrightarrow \perp\!\!\!\perp(Y, X, Z).$$

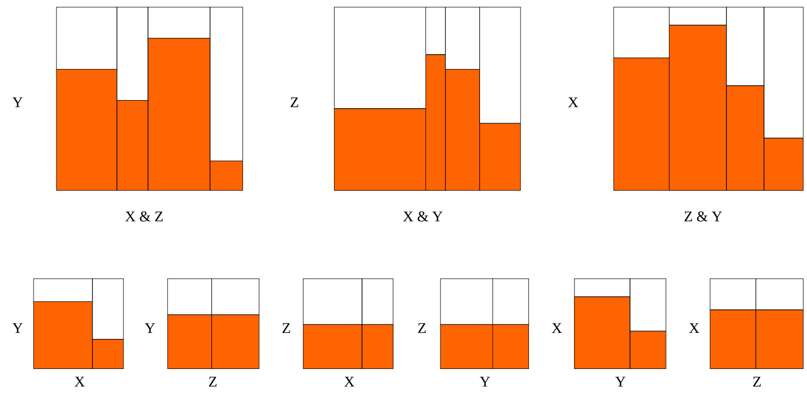


Figure 13: No flats. Two marginal independencies. No other independence

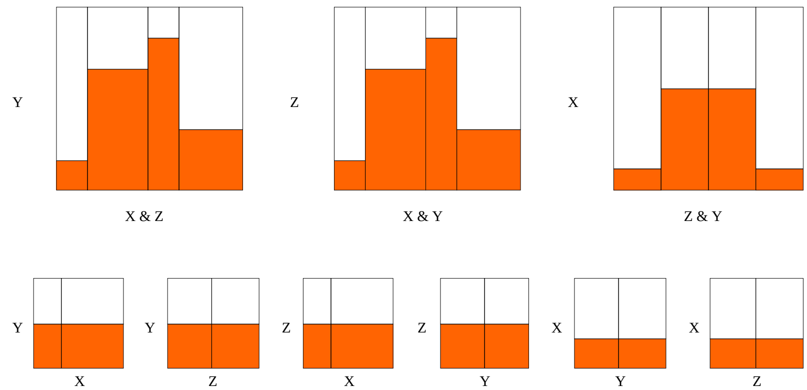


Figure 14: No flats. Three marginal independencies. No other independence

is perhaps the most counter-intuitive. The eikosograms of Figure 14 not only demonstrate the possibility but also give some indication as to how it can occur.

4 Statistical models

A number of different statistical models have been suggested to meaningfully describe the joint distribution of categorical random variables. Two classes of models which are currently advocated are graph-based models, especially graphical models, and log-linear models. The first of these is motivated largely by appeal to a graph theoretical representation of the joint distribution; the second is motivated by analogy to the analysis of variance style models which grew out of classic experimental design.

4.1 Graph-based models

In graph-based models, interest lies predominantly in determination of independencies that exist between the variables involved. Intuition is derived from a graph where variables are represented as nodes and dependencies as the arcs which link the nodes. – presence of an arc indicates a given kind of pairwise dependence, absence the corresponding pairwise independence.

Figure 15 shows all such graphs for three variables (modulo permutation of the variables), each a different model for the joint distribution. The visual emphasis which these graphs give to probabilistic dependence over independence

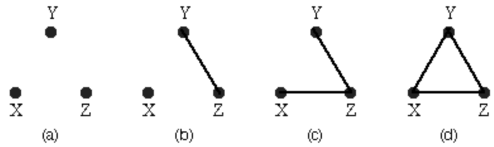


Figure 15: *Dependence graphs*: A summary of the joint distribution of the variables which visually emphasizes some kind of pairwise (possibly conditional) dependency. Nodes are variables and an edge between variables a dependency; no edge indicates an independence of some kind.

suggests the name *dependence graphs*. This visual emphasis has the models corresponding to Figures 15(a)-(d) appear to increase in complexity from left to right as more arcs are added to the diagram.

When one considers that independence between two variables is an assertion which requires testing, then left to right the models actually *decrease* in the number of assertions made about the joint distribution. The increased simplicity of the model dependence structure is had only by increased complexity of the independence structure imposed.

The *independence graphs* of Figure 16 instead match visual complexity to that of the independence assertions

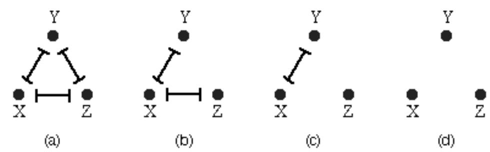


Figure 16: *Independence graphs:* The complement of those of Figure 15. Here the arcs do not quite connect and are blunted as if keeping the nodes/variables apart to emphasize the barrier of independence rather than the join of dependence.

made by the model they represent. The arcs are shown blunted and not quite reaching the nodes so as to suggest independence denying or barring the connection of dependence; for mnemonic convenience the blunted arcs will be called '`I`-bars'. Figures 16 (a)-(d) describe the same models as those of Figures 15(a)-(d), respectively. Although graph-based modellers at present make use only of dependence graphs, both visual representations need to be kept in mind when modelling: the dependence graph Figure 15 indicates the simplicity of the model, while its counterpart Figure 16 shows the complexity of the model's assumptions. Whenever one graph is sparsely connected, it will be the easier of the two to comprehend and suggests that one graph might be preferable to the other in some situations.

Using the graphs of Figure 15, it is easy to visually associate complete independence, $\perp\!\!\!\perp(Y, X, Z)$, with (a) and complete dependence, $\bowtie(Y, X, Z)$, with (d); but this would be a mistake. What can be asserted, or not, depends upon the kind of dependence which an arc in Figure 15 represents. Given the dependence type of the arc, the eikosograms of the last section can now be put to good use to better understand what each model in Figure 15 has to say about the underlying distribution.

4.1.1 Conditional dependence graphs

The original (Darroch, et al, 1980) and most common (e.g. Whittaker, 1990, Lauritzen, 1996, Cox and Wermuth, 1996) graphical model is that where the presence of an arc between two nodes in any graph in Figure 15 means that the two variables are *conditionally dependent given all other variables in the graph*, the absence means conditional independence. These models were first introduced (Darroch, et al, 1980) as a subset of the hierarchical log-linear models and have seen much use in artificial intelligence programs – first for the propagation of uncertainty in expert systems (e.g. Lauritzen and Spiegelhalter, 1988) in sparsely connected graphs and now for some knowledge representation applications in data mining.

The dependence graph becomes a *conditional dependence graph* and the independence graph a *conditional independence graph*. The former has traditionally been called a conditional *in*dependence, rather than dependence, graph (e.g. Whittaker, 1990) largely because only this choice of display has been used for graph-based modelling. Given its visual emphasis, the traditional naming seems misleading and better suited to graph representations like those of Figure 16. Cox and Wermuth (1996, p. 30) use the name *concentration graph* although this too seems an unfortunate choice based as it is on appeal to the concentration matrix (inverse covariance matrix) of a multivariate normal distribution for its conditional independence interpretations – absent the multivariate normal joint distribution and the concentration matrix interpretation becomes focused on the presence or absence of *linear* conditional dependence for which the term 'concentration graph' might better be reserved (see Cox and Wermuth, 1996, pp. 63-69).

Because flats are the visual manifestation of independence in an eikosogram, the conditional independence graphs are easily matched to characteristics of an eikosogram. Each `I`-bar of a conditional independence graph in Figure 16 marks the presence of a 2×2 flat in each of two three-way eikosograms, namely those having one of the two variables involved as the vertical variable. A 'V' configuration, as for example Y to X to Z in Figure 16(b), indicates a 4-flat in the corresponding three-way eikosogram when X, the variable at the point of the 'V', is the vertical variable of the eikosogram. As a consequence marginal independence will hold as well.

4.1.2 Marginal dependence graphs

For model development, it can be argued that marginal dependence or independence of two variables is relatively easy to think about because the other variables are essentially ignored by the marginalization. Certainly marginal pairwise independence helps model interpretation afterwards. In either case, having the arcs of Figure 15 indicate marginal dependence (i.e. unconditional dependence) provides a help to understanding the joint distribution of the variables.

With this interpretation, each dependence graph of Figure 15 is now a *marginal dependence graph* and each independence graph of Figure 16 a *marginal independence graph*. Cox and Wermuth (1996, p. 30) call the former a *covariance graph* even though the motivation for associating zero covariance with independence comes from their identity in the case of multivariate normal random variables. Again, the idea of covariance is much too restrictive to meaningfully describe a graph whose arcs represent marginal dependencies in an arbitrary joint distribution; better to use the longer but accurate 'marginal dependence graph' and reserve covariance graph for those situations when linear dependence (or not) is being shown.

The `I`-bars of the marginal independence graph now mark the fact that either marginal two-way eikosogram for the pair involved is flat; it says nothing about the flatness of any three-way eikosogram. The existence of a 'V' in the marginal independence diagram is not as strong as in a conditional independence diagram and says only that both sets of two-way eikosograms are flat.

Because the arcs (or `I`-bars) of a marginal dependence (independence) graph are determined entirely from the two variables involved, unlike their conditional counterparts, marginal realtionship graphs can be built up incrementally adding variables as they become of interest or available. Introducing a new variable in a conditional graph affects every pairwise relationship in the graph – all arcs (and `I`-bars) would need to be re-determined.

4.1.3 Inability of either graph-based model to distinguish cases.

As has already been noted, neither graph is intended to show independence relations involving single events. An example of some practical import would be the conditional independence of variables for a *particular value* of a third, e.g. $Y \perp\!\!\!\perp X | \{Z = \mathsf{n}\}$ as was built into the eikosograms of Figure 12. If for example Y represents survival, X a medical treatment and Z the patient's sex, then it would be important to know, for example, that the treatment was ineffective for females, however effective it might have been for males. While visually clear in an eikosogram (e.g. Figure 12) this would be completely obscured in a conditional dependence graph. Independencies which occur only for particular values of the conditioning variable fall outside the domain of these models.

When examing the various possibilities of independence amongst three variables which were outlined in Section 3, the two types of graph-based models match cases with varying success. Figure 17 shows the difficulties in going from either (conditional or marginal) graph-based model to the corresponding histograms.

Reading across the third row, we see that the conditional independence graphs successfully match only the first two cases corresponding to the eikosograms of Figures 7 and 8. While there is a match on Cases 3 as well, this corresponds to two separate independence structures which are indistinguishable from one another via the graphical model. The match is especially bad for the fourth graph. There one graph covers four separate independence structures which expands to eight when all possible combinations of the three variables is considered!

Independence graph				
Dependence graph				
Conditional	Case 1 Figure 7	Case 2 Figure 8	Case 3 Figures 9 and 10	Case 4 Figures 11, 12, 13, and 14
Marginal	Cases 1 and 4 Figures 7 and 14	Cases 2 and 4 Figures 8 and 13	Cases 3 and 4 Figures 10 and 12	Cases 3 and 4 Figures 9 and 11

Figure 17: Matching graph-based models to eikosogram cases.

For the marginal graphs, the match up is even worse; each graph corresponds to at least two possible sets of independence structures.

4.1.4 Combined marginal and conditional dependence graphs

Cox and Wermuth (1996) suggest using dashed lines between variables to indicate marginal pairwise dependence and solid lines to indicate conditional dependence. However they use these different line styles only to distinguish the marginal dependence graph from the conditional dependence graph even though the different line styles would permit both graphs to be merged into one. For three variables the combined graph will capture and display every possible variable independence structure.

The eikosograms of Figures 7 to 14 are now each matched by either one of two combined graph representations – one being the *combined dependence graphs* of Figure 18, the other being the *combined independence graphs* of Figure

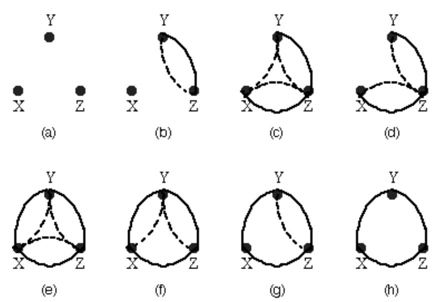


Figure 18: *Combined dependence graphs:* First row: (a) Case 1 of no dependencies as in the eikosgrams of Figure 7, (b) Case 2 with one marginal and corresponding conditional dependency as in Figure 8, (c) Case 3 as in Figure 9, (d) Case 3 but as in Figure 10 where Z must have more than two categories; Second row: The four possibilities for all conditional dependencies of Case 4, (e) All marginal and conditional dependencies as in Figure 11, (f) Two marginal dependencies as in Figure 12, (g) One marginal dependency as in Figure 13, (h) No marginal dependencies as in Figure 14.

19. Dashed arcs (`I`-bars) represent marginal dependence (independence), solid arcs (`I`-bars) conditional dependence (independence); arcs are now curved with the marginal ones on the inside of the triangle.

As before, the graphs of Figure 18 and Figure 19 are complementary.

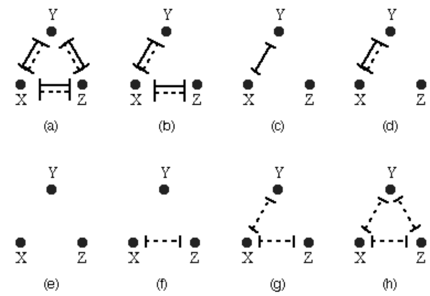


Figure 19: *Combined independence graphs:* First row: (a) Case 1, All marginal and conditional independencies as in the eikosgrams of Figure 7, (b) Case 2 with two marginal and corresponding conditional independency pairings as in Figure 8, (c) Case 3 with a single conditional independency as in Figure 9, (d) Case 3 with both one conditional and one marginal independency as in Figure 10, Z must have more than two categories; Second row: The four possibilities for no conditional independence of Case 4, (e) No additional marginal independencies as in Figure 11, (f) One marginal independency as in Figure 12, (g) Two marginal independencies as in Figure 13, (h) Three marginal independencies as in Figure 14.

4.2 Log-linear models

Log linear models are perhaps the most common statistical model for contingency tables. The model is expressed in terms of the natural logarithm of the joint probability of all variables and u-terms representing anova style additive effects (overall, main, two-way and three-way interaction). For three categorical variables the model has $p_{ijk} = Pr(Y = y_i, X = x_j, Z = z_k)$ for $i = 1, \ldots I$, $j = 1, \ldots J$, and $k = 1, \ldots K$, and

$$log(p_{ijk}) = u + u_{Y(i)} + u_{X(j)} + u_{Z(k)} + u_{YX(ij)} + u_{YZ(ik)} + u_{XZ(jk)} + u_{YXZ(ijk)}.$$

Being overparameterized wrt to the u-terms, the ‘usual constraints’ imposed on the u-terms are that the sum of any u-term over any one of its indices is zero (e.g. see Fienberg, 1977).

Hierarchical log-linear models are those for which a zero low-order interaction (or main) term implies that all higher order terms involving the same variables are also zero. For example, the log-linear model having $u_{YX(ij)} = 0$ for all i and j but $u_{YXZ(ijk)} \neq 0$ for some i, j and k is not hierarchical. Non-hierarchical models are generally difficult to interpret and are not used for that reason (computationally, they can be more difficult to fit as well).

Hierarchical log-linear models include the graphical models (i.e. conditional dependence graph based models) as a proper subset. Graphical models are identified with setting to zero different two way u-terms (and all higher, being hierarchical); the result is conditional independence between those variables which define the two way interactions (given all others in the model). A model which has all three way and higher order terms zero but no two way terms zero is an example of a hierarchical non-graphical log-linear model.

Conditional independence of two variables given all other variables in the model and complete independence of variables can be expressed by setting relevant u-terms to zero. Because marginalization requires summation over $exp(log(p_{ijk}))$, there is no similarly simple relationship between setting u-terms to zero and marginal independence (absent conditional independence); the same is true for conditional independence given a proper subset of the remaining variables.

Hierarchical models which are non-graphical can give insight into the nature of the dependence as opposed to the independence. For three variables the only non-graphical hierarchical model has $u_{YXZ(ijk)} = 0$ for all i, j and k but no other u-term is everywhere zero. While asserting nothing about independence, the zero three way interaction ensures some consistency in the conditional dependencies. Figure 20 shows three examples where $u_{yxz(ijk)} = 0$. In the $2 \times 2 \times 2$ case, the conditional dependencies have the same direction whatever the value of the third variable. The third example is graphical in that a conditional independence also exists.

Log-linear models are also capable of expressing a conditional independence for particular events. For three variables, if say $u_{YX(ij)} = u_{YXZ(ijk)}$ only for $k = 1$ say, then the $Y \perp\!\!\!\perp X | \{Z = z_1\}$. Only if equality holds for all k, in which case the usual constraints force equality to zero as well, will $Y \perp\!\!\!\perp X | Z$.

Like the graphical models, a single log-linear model (hierarchical or not) does not distinguish the different possibilities of Cases 3 and 4. Unlike the individual graph-based methods, log-linear models can handle conditional

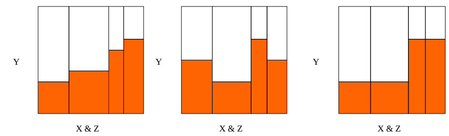


Figure 20: Three examples where $u_{YXZ(ijk)} = 0$ for all i, j, k.

independence given single events and will not be mute about the nature of the dependencies.

5 Theorems for three variables from doubly combined graphs

Several theoretical results are either directly apparent from the eikosograms or can be proven by algebraically matching the relevant visual characteristics. The combined dependence and independence graphs for three variables suggest a number of fundamental relationships which underlie the graphs of Figures 18 and 19. These two visual displays, eikosograms and graphs, can now be worked together to prove and illustrate a number of theorems on the independence relations between three variables.

The complementarity of the dependence and independence graphs means that the two can be put together on the same graph to produce a complete graph showing every dependence as an arc and every independence as an `I`-bar. Using different line styles, solid for conditional and dashed for marginal, allows the two complete graphs (one dashed, one solid) to be put together as well. Figure 21 shows the *doubly combined graphs* which result from putting together

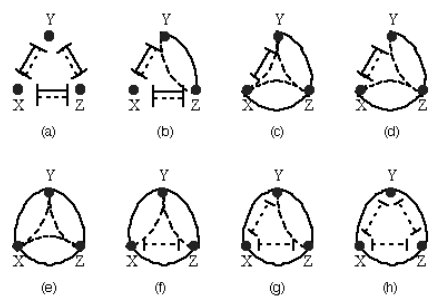


Figure 21: *Doubly combined graphs:* Putting Figures 18 and 19 together produces doubly complete graphs which show every dependence and every independence explicitly.

the graphs of Figures 18 and 19. The result is a doubly complete graph which expresses *all* of the independence and dependence structures *explicitly* in the one display.

Each graph in Figure 21 provides an easily interpreted graph description for all the eikosograms of that case from Section 3. Independencies, being straight-lined `I`-bars in the graph suggest corresponding flats in an eikosogram; dependencies as curves in the graph visually suggest non-flats in an eikosogram. Three way eikosograms (i.e. conditional ones) are built from following the solid lines, two way (i.e. marginal eikosgrams) follow the dashed lines.

Knowing that the goal is to produce a complete doubly combined graph, the components can be used to assert what is known about the independence relations. For example, the incomplete graph of Figure 22(a) is interpreted as asserting only that Y and Z are conditionally dependent given X. Because no other relationship (dependence or independence) appears in Figure 22(a), no more than what is seen has been asserted – i.e. nothing is said for example about the marginal relationship of Y and Z nor about the relationship, conditional or marginal, between Y and X or between X and Z. Similarly, Figure 22(b) asserts only that $X \perp\!\!\!\perp Z$ and expresses no information about any other relationship; any one of the graphs (a), (b), (f), (g), or (h) of Figure 21 would correspond to Figure 22(b).

This interpretation provides a means of visually expressing, and even proving, results on the independence of three

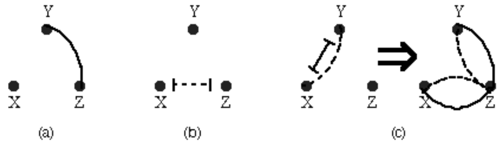


Figure 22: *Graphically asserting pairwise relations.* No connection/I-bar means no statement about the pairwise relationship while connections and I-bars mean the same as before. (a) $Y \not\perp\!\!\!\perp Z|X$, other (in)dependencies may or may not exist, (b) $X \perp\!\!\!\perp Z$, other (in)dependencies may or may not exist, and (c) A graphical statement of the theorem $Y \perp\!\!\!\perp X|Z$ and $Y \not\perp\!\!\!\perp X$ is sufficient to produce the structure seen in Case 3, Figures 9 and 21(c).

variables. Figure 22(c), for example, provides a visual statement of a theorem:

$$Y \perp\!\!\!\perp X|Z \quad \text{and} \quad Y \not\perp\!\!\!\perp X \Longrightarrow Z \not\perp\!\!\!\perp Y, \quad Z \not\perp\!\!\!\perp X, \quad Z \not\perp\!\!\!\perp Y|X, \quad \text{and} \quad Z \not\perp\!\!\!\perp X|Y.$$

5.1 Flat water theorems

Theorems such as this can be built up from smaller, less complex results. The most fundamental of these are what might mnemonically be called "flat water" theorems because their proof follows directly from the water barrier metaphor when the water is the same level, or flat, across the barrier. These are summarized graphically in Figure 5.1.

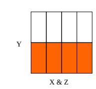
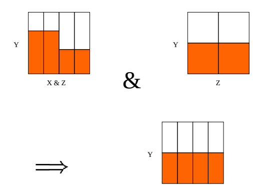
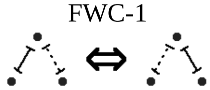
Label	Graph representation	Symbolically	Proof
FW1		$Y \perp\!\!\!\perp X\|Z$ and $Y \perp\!\!\!\perp Z\|X$ $\Longrightarrow Y \perp\!\!\!\perp X$ and $Y \perp\!\!\!\perp Z$.	
FW2		$Y \perp\!\!\!\perp X\|Z$ and $Y \perp\!\!\!\perp Z$ $\Longrightarrow Y \perp\!\!\!\perp Z\|X$.	

Figure 23: Flat water theorems.

The proofs of these two theorems follow directly from examining the relevant eikosograms (as shown in the third column of Figure 5.1) and applying the water container metaphor for marginalization - beginning with flat water can only yield flat water. These results hold however many categories each of the three variables might have.

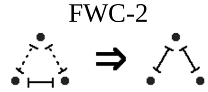
The leftmost graphical relation of Figure 24 (i.e. FWC-1) shows a symmetry theorem which can be proven simply

FWC-1



$Y \perp\!\!\!\perp X|Z$ and $Y \perp\!\!\!\perp Z \iff Y \perp\!\!\!\perp X$ and $Y \perp\!\!\!\perp Z|X$

FWC-2



$Y \perp\!\!\!\perp X$, $Y \perp\!\!\!\perp Z$ and $X \perp\!\!\!\perp Z|Y$ $\Longrightarrow$ $Y \perp\!\!\!\perp X|Z$ and $Y \perp\!\!\!\perp Z|X$ (or $\perp\!\!\!\perp(Y, X, Z)$)

Figure 24: Immediate corollaries of the flat-water theorems.

by appeal to the graphical statements of the flat water theorems. The right figure is a rotational symmetry theorem proven simply by repeated application of FW2. Together these show that complete independence between three variables holds iff a complete graph whose edges are `I`-bars can be formed, of which at least one edge is a conditional `I`-bar.

Some other simple but fundamental results can also be proven entirely by appealing to the graphs. These are summarized in Figure 25. Knowing that a doubly combined graph cannot have both an `I`-bar and an arc of the same

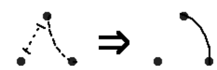
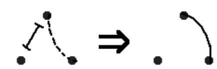
Label	Graph representation	Symbolically
FWC-3		$Y \perp\!\!\!\perp X\|Z$ and $Y \not\perp\!\!\!\perp Z\|X$ $\Longrightarrow Y \not\perp\!\!\!\perp X$
FWC-4		$Y \perp\!\!\!\perp X$ and $Y \not\perp\!\!\!\perp Z$ $\Longrightarrow Y \not\perp\!\!\!\perp X\|Z$
FWC-5		$Y \perp\!\!\!\perp X\|Z$ and $Y \not\perp\!\!\!\perp Z$ $\Longrightarrow Y \not\perp\!\!\!\perp X\|Z$
FWC-6		$Y \perp\!\!\!\perp X\|Z$ and $Y \not\perp\!\!\!\perp X$ $\Longrightarrow Y \not\perp\!\!\!\perp Z\|X$ and $Y \not\perp\!\!\!\perp Z$
FWC-7		$Y \not\perp\!\!\!\perp X\|Z$ and $Y \perp\!\!\!\perp X$ $\Longrightarrow Y \not\perp\!\!\!\perp Z\|X$

Figure 25: Corollaries of the flat-water theorems via proof by contradiction.

type (conditional or marginal; solid or dashed) between two nodes, these results follow from the flat water theorems via proof by contradiction.

For example, FWC-3 follows by assuming to the contrary that a marginal independence is the consequence. But FW-2 asserts that the conditional dependence seen in the premise of FWC-3 must be therefore be a conditional independence – a contradiction.

The corollary FWC-4 is of especial statistical interest. Label the nodes with Z at the bottom left, X at the top, and Y at the bottom right and think of X as an experimental treatment and Y as a response and Z being of little interest. Then if Z is known, by blocking on its values, we can assure that the treatment is marginally independent of the blocking variable Z. Consequently observing a marginal dependence between Y and X means that a dependence will be observed conditionally as well. If Z is not known, then the act of randomly allocating the treatment ensures that X and Z are marginally independent and so observing the marginal dependence (which is all that is possible to observe, since Z is not known) guarantees that a conditional independence holds as well. Note that seeing a marginal independence between response and treatment guarantees nothing about the existence of a conditional independence. More can be said on the statistical interpretation of these theorems and they provide interesting starting points for discussion.

From these few fundamental theorems, a number of results follow. The theorem of Figure 22 follows from twice applying FWC-6. Figure 26 illustrates a few other theorems similarly proved. For example, Figure 26(a) indicates that although three variables are pairwise independent, should any two be conditionally dependent given the third, then every pair is conditionally dependent given the remaining variable. The remaining Figures 26(b)-(d) can be similarly interpreted.

Together these consequences of the flatwater theorems are sufficient to enumerate almost all of the independence structures seen in the eikosograms of Section 3 and summarized in the doubly complete graphs of Figure 21. It only

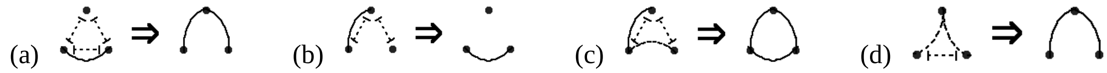


Figure 26: Some further consequences of the flat-water theorems provable via the graphs.

remains to show that the independence structure given by Figure 21 (d) is not possible when the conditioning variable is binary.

5.2 Rubber wishbone results when the connecting variable is binary

Binary variables are ubiquitous – male and female, treatment-A and treatment-B, and so on. In the case where at least one variable is binary (the other two can have any number of categories ≥ 2), a number of other results can be shown to hold which do not hold otherwise. The results require that the dependence/independence structure is 'wishbone shaped' with the binary variate appearing at the join of the wishbone.

The fundamental result is a *rubber wishbone* theorem which is stated graphically and symbolically in Figure 27. Its proof is given in the Appendix. The graphical representation of the conditions of the theorem (left hand of the

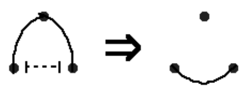
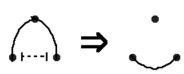
Graph representation	Symbolically	Proof
	Y binary: $Y \not\perp\!\!\!\perp X\|Z, Y \not\perp\!\!\!\perp Z\|X$ and $Z \perp\!\!\!\perp X$ $\implies Z \not\perp\!\!\!\perp X\|Y$.	See Appendix.

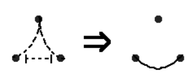
Figure 27: *Rubber wishbone* theorem holds for binary Y. It does *not* hold if Y has more than two values.

graphical representation in Figure 27) look like a wishbone where each dependency is identified with a bone and the independency appears to be pushing the bones apart.

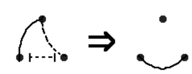
In the case of Figure 27 the dependencies are conditional and the independency is unconditional but this need not be the case. Figure 28 shows six results which are equivalent in the sense that if any one of these theorems hold, then



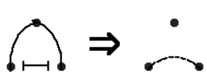
$Y \not\perp\!\!\!\perp X|Z, Y \not\perp\!\!\!\perp Z|X$, and $Z \perp\!\!\!\perp X$
$\implies Z \not\perp\!\!\!\perp X|Y$.



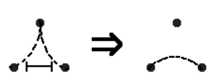
$Y \not\perp\!\!\!\perp X, Y \not\perp\!\!\!\perp Z$ and $Z \perp\!\!\!\perp X$
$\implies Z \not\perp\!\!\!\perp X|Y$.



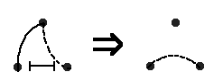
$Y \not\perp\!\!\!\perp X|Z, Y \not\perp\!\!\!\perp Z$ and $Z \perp\!\!\!\perp X$
$\implies Z \not\perp\!\!\!\perp X|Y$.



$Y \not\perp\!\!\!\perp X|Z, Y \not\perp\!\!\!\perp Z|X$, and $Z \perp\!\!\!\perp X|Y$
$\implies Z \not\perp\!\!\!\perp X$.



$Y \not\perp\!\!\!\perp X, Y \not\perp\!\!\!\perp Z$ and $Z \perp\!\!\!\perp X|Y$
$\implies Z \not\perp\!\!\!\perp X$.



$Y \not\perp\!\!\!\perp X|Z, Y \not\perp\!\!\!\perp Z$ and $Z \perp\!\!\!\perp X|Y$
$\implies Z \not\perp\!\!\!\perp X$.

Figure 28: The rubber wishbone: these several variations of rubber wishbone results (binary Y) are equivalent in the sense that if any one of these hold, they all hold. The equivalencies are easily proved graphically using flat-water results.

they must all hold. Each can be proved graphically to follow from any one of the others.

The name of the theorem is meant to present an image of its content. Choose any of the equivalent statements in Figure 28. The nature of the theorem is that: if a wishbone shape of dependencies holds on three variables and the variable at the join is binary, then this wishbone cannot be 'pulled apart' in the sense of having both conditional and marginal independences *simultaneously* hold between the two end variables. If one pulls at the wishbone's ends by inserting a conditional independence, then a marginal dependence brings the two ends back together; pull at the ends by inserting a marginal independence and a conditional dependence brings the two ends back together – it is as if the wishbone were made of rubber. This holds no matter which dependencies (conditional or marginal) define the wishbone shape. But, if the joining variable is *not binary*, then the wishbone is not rubber and can break in the sense that both kinds of independences can exist simultaneously between the two end variables.

5.2.1 Some inferential consequences

The inferential consequences of the rubber wishbone results can be subtle and are perhaps best understood when the variables are meaningful. A simple but general case of some import is that seen in Figure 29. There the explanatory

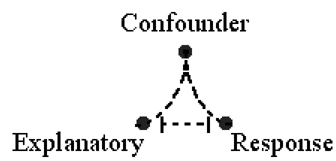


Figure 29: Wishbone configuration between an explanatory variable of interest, a response variable and some confounding variable marginally associated with each of the other variables.

variable of interest is marginally independent of the response as might be observed when considering the variables in some study. However, suppose there exists a confounding (sometimes sinisterly called 'lurking') variable which is marginally associated with each of the response and the explanatory variable. The result is a wishbone shape.

Scientifically, it can be desirable that a marginal independence observed between the explanatory variable and the response also hold when a third variable associated with both is taken into account – that is, conditionally on the third the explanatory and response variate remain independent. The reason is that this introduces some simplification as the explanatory variable is usually then discounted in terms of its ability to explain the response; other explanatory variables are pursued instead. Of course when this is not the case, further investigation of the relationship between explanatory variable and response conditional on the third variable is pursued.

If the confounding variable has more than two possible categories, then the conditional independence might or might not hold between the explanatory and the response given the confounder. However, should the confounding variable be binary, then the rubber wishbone dictates that the explanatory and response variables are necessarily conditionally dependent.

A scenario where this consequence might be of concern is as follows. Suppose that the wishbone configuration is obtained with a multi-category (> 2) background variable and further that the response and explanatory variable are conditionally independent given the background variable. Scientifically, then, some simplification seems possible. However, suppose that we combine categories of the confounding variable until the confounding variable is binary and the wishbone structure is preserved (this is possible). Because the confounder is now binary, the explanatory and response variables must be conditionally dependent and scientifically merit further investigation. Where independence existed between the explanatory variable and the response both marginally and conditionally given the values of the confounder the conditional independence can be destroyed simply by reducing the number of categories of the confounder to two and the scientific investigation possibly turn in a different direction.

While combining categories mathematically changes the variable from what it was, it might not cause investigators to substantively change its meaning in the context of a study. Imagine a study for each of the fictitious examples in Figure 30.

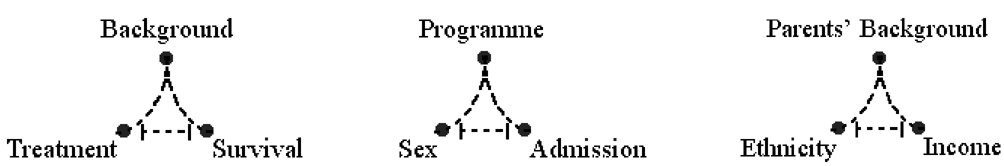


Figure 30: Fictitious examples to consider some inferential consequences of the rubber wishbone theorem.

In the first example, a background variable (e.g. features of the patient's prior hospitalization record) on patients in a medical study is seen to exhibit the wishbone structure. If the background variable has more than two categories, then it may be that the conditionally on this variable the survival probabilities are independent of the different treatments given in the study. The disturbing thing is that by reducing the background variable's categories to two (preserving the wishbone structure) will result in a conditional dependence which suggests the treatments might be effective (or not) depending on which of two categories of background variable the patient belongs to. More time and effort will then be spent on determining why this is the case when it may in fact be entirely an artifact of the rubber wishbone theorem.

In an experimental study, this problem can be avoided. Randomly allocating treatment values to patients ensures that any background variable whose values are determined (though possibly unknown) prior to treatment are distributed independently of the treatment variable and so the wishbone structure cannot occur. The difficulties remain where random allocation is not possible, or not done.

The two remaining examples provide socio-economic contexts where studies are more likely to be observational than experimental. The first imagines that there is no relationship between the sex of an applicant and their consequent admission to university (a twist on the now classic Berkeley Graduate School admissions version of Simpson's paradox). If, however the programme to which the application is made is considered then a wishbone structure can occur and the consequent difficulty of the conditional relationship between sex and admission depending on the number of programmes defined.

The last context supposes that a person's income category does not depend on their ethnic background. Some feature of their parents' background (e.g. citizenship, education, ethnicity, urban/rural, etc.) could produce a wishbone structure and consequently the possibility of introducing a conditional relationship between ethnicity and income simply by forcing the parents background variable to be binary.

As these examples illustrate, the rubber wishbone results can be avoided by good statistical practice. Variables whose values are known at the plan stage of the study can be made distributionally independent of the explanatory (or treatment) variable by matching study units and so avoid the wishbone structure. Random allocation, if available, will introduce the same independence between the treatment and unknown (but pre-determined) confounders. To this practice it must be added that care needs to be taken in the construction of the number of categories which a potential confounder takes and interpretations based on binary confounder variables be made with caution.

6 Concluding remarks

Exploring the independence possibilities of three variables is an important topic which merits solid treatment in introductory probability courses. Eikosograms can be used to assist that exploration and to help ground the understanding. They provide a concrete basis for discussion without the necessity of a mathematical treatment. Other independence structures (e.g. Markov process) could be given visual display via eikosograms as well.

Dawid's notation (slightly extended to cover events and n-ary arguments) is well suited to describe the independence features visible on the eikosograms. The flat water graphs which fall out of the flat water eikosograms provide visual tools for formal reasoning about independence structures – again no symbolic representation is needed for all but one proof. From the flat water theorems, the corollaries are so easily derived that they need never be committed to memory. The nature of the dependence and any independence involving only events, as opposed to variables, will be well represented by eikosograms but not at all by the graphs.

A deep understanding of these possibilities leads to deeper understanding of the structure of statistical models for categorical data. Graphical models having sparse dependence graphs appear visually benign but contain a great deal of independence assertions which are visually hidden; these are given explicit emphasis in independence graphs. Combining both types of graphs gives more accurate visual description of the model involved. Lower order in/dependencies (e.g. marginal) will be obscured in graphical and log-linear models (and in high order eikosograms, absent the lower order ones).

For three variables the two kinds of in/dependencies given by solid and dashed edges are sufficient to graphically represent all possibilities. For more than three variables, they will not be sufficient and the flat water graphs will not apply without extension. It should be noted however that all of the three variable results hold when all are understood to be conditional on any other set of variables (e.g. U, W, ...) and so retain value in more complex situations. There are extensions of the flat theorems that correspond to those of graphical models and isolated groups of nodes. In the case of four and more variables, the notation of complete independence and complete dependence, marginally and conditionally, could be used to simplify or possibly focus the description.

A deep understanding of even the case of three variables provides a base for understanding the consequences on statistical inference of certain design decisions as noted in Section 5. This understanding could be put to good use in statistics courses which followed the probability course.

Acknowledgements

I am grateful to my colleagues Profs. Winston Cherry and Mary Thompson at Waterloo: to Winston Cherry for discussions on eikosograms leading to this paper and to Mary Thompson for her careful reading and comments on an earlier draft which helped improve the presentation.

Appendix

Parameter values for the figures

For three binary variables, the eikosograms are simply parameterized by eight letters $a, b, \ldots, h$ as shown in the diagram for Y vs X & Z in canonical form as shown in Figure 31. Following the canonical layout we have: $Pr(X =$

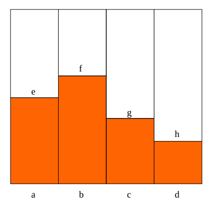


Figure 31: Parameter values for Y vs X & Z in canonical form: bar widths a b c and d; bar heights e f g and h.

$\mathsf{y}, Z = \mathsf{y}) = a$, $Pr(X = \mathsf{n}, Z = \mathsf{y}) = b$, etc. and $Pr(Y = \mathsf{y}|X = \mathsf{y}, Z = \mathsf{y}) = e$, $Pr(Y = \mathsf{y}|X = \mathsf{n}, Z = \mathsf{y}) = f$, etc.

The values for the Figures in the paper (where not available on the Figure itself) are given in Table 1.

Figure	a	b	c	d	e	f	g	h
5(a)	2/7	8/35	1/7	12/35	4/5	1/2	3/10	7/10
5(b)	2/7	8/35	1/7	12/35	4/5	1/2	3/5	1/5
5(c)	2/7	8/35	1/7	12/35	4/5	1/2	4/5	1/2
6	2/7	8/35	1/7	12/35	7/10	7/10	7/10	7/10
7	10/33	4/11	5/33	2/11	7/10	7/10	7/10	7/10
8	1/3	1/6	1/3	1/6	7/10	7/10	3/10	3/10
9	2/9	1/9	2/9	4/9	2/3	2/3	1/6	1/6
11	1/7	1/7	3/7	2/7	1/3	2/3	1/4	1/6
12	1/4	1/4	1/4	1/4	3/4	1/2	1/4	1/4
13	1/3	1/6	1/3	1/6	2/3	1/2	5/6	1/6
14	1/6	1/3	1/6	1/3	1/6	2/3	5/6	1/3

Table 1: Parameter values for the figures: bar widths a b c and d; bar heights e f g and h.

These parameters can also be expressed in terms of the usual contingency table notation:

$$a = p_{011} + p_{111} \qquad e = p_{111}/a \qquad \text{or equivalently} \qquad p_{111} = a \times e \qquad p_{011} = a \times (1 - e)$$

$$\begin{array}{llll} b = p_{001} + p_{101} & f = p_{101}/b & p_{101} = b \times f & p_{001} = b \times (1-f) \\ c = p_{010} + p_{110} & g = p_{110}/c & p_{110} = c \times g & p_{010} = c \times (1-g) \\ d = p_{000} + p_{100} & h = p_{100}/d & p_{100} = d \times h & p_{000} = d \times (1-h) \end{array}$$

Rubber wishbone theorem

Theorem: When Y is a binary random variable, $Y \not\perp\!\!\!\perp X|Z, Y \not\perp\!\!\!\perp Z|X$ and $X \perp\!\!\!\perp Z$ together imply $X \not\perp\!\!\!\perp Z|Y$.
Proof: It will be more convenient in the proof to use the following notation notation:

$$a_{jk} = Pr(X = x_j, Z = z_k) \quad \text{and} \quad e_{ijk} = Pr(Y = y_i | X = x_j, Z = z_k)$$

for all $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$ representing the distinct values of the three variables ($I = 2$ and $J, K > 1$). These values give the dimensions of the Y vs $X \& Z$ eikosogram – a_{jk}s provide the bar widths and e_{ijk}s the heights. All probabilities will be taken to be strictly between 0 and 1.

$Y \not\perp\!\!\!\perp X|Z$ means that for at least one j_*, j_{**} and k_* with $j_* \neq j_{**}$, we have

$$e_{ij_*k_*} \neq e_{ij_{**}k_*} \quad \text{for } i = 1, 2. \tag{4}$$

Similarly, $Y \not\perp\!\!\!\perp Z|X$ means that for at least one j', k' and k'' with $k' \neq k''$, we have

$$e_{ij'k'} \neq e_{ij'k''} \quad \text{for } i = 1, 2. \tag{5}$$

The third premise, $X \perp\!\!\!\perp Z$ implies that

$$\begin{aligned} Pr(X = x_j | Z = z_{k_1}) &= Pr(X = x_j | Z = z_{k_2}) \quad \text{for all} \quad j, k_1, k_2 \\ \text{or} \quad \frac{a_{jk_1}}{a_{+k_1}} &= \frac{a_{jk_2}}{a_{+k_2}} \quad \text{for all} \quad j, k_1, k_2 \end{aligned} \tag{6}$$

where a '$+$' subscript means summation over all values of that subscript.

The theorem is now proved by contradiction. Assuming that $X \perp\!\!\!\perp Z|Y$ holds as well implies that for all i, j_1, k_1 and k_2

$$Pr(X = x_{j_1} | Y = y_i, Z = z_{k_1}) = Pr(X = x_{j_1} | Y = y_i, Z = z_{k_2})$$

$$\begin{aligned} \iff \quad & \frac{a_{j_1k_1} e_{ij_1k_1}}{\sum_j e_{ijk_1} a_{jk_1}} = \frac{a_{j_1k_2} e_{ij_1k_2}}{\sum_j e_{ijk_2} a_{jk_2}} \\ \iff \quad & \frac{e_{ij_1k_1}}{e_{ij_1k_2}} = \frac{a_{j_1k_2} \sum_j e_{ijk_1} a_{jk_1}}{a_{j_1k_1} \sum_j e_{ijk_2} a_{jk_2}} \end{aligned} \tag{7}$$

Together with equation (6) this means that

$$\frac{e_{ij_1k_1}}{e_{ij_1k_2}} = \frac{a_{+k_2} \sum_j e_{ijk_1} a_{jk_1}}{a_{+k_1} \sum_j e_{ijk_2} a_{jk_2}}$$

the right hand side of which no longer depends on the value of j_1. This means that for all i, j_1, j_2, k_1, and k_2 we have

$$\frac{e_{ij_1k_1}}{e_{ij_1k_2}} = \frac{e_{ij_2k_1}}{e_{ij_2k_2}} \quad \text{or equivalently} \quad \frac{e_{ij_1k_1}}{e_{ij_2k_1}} = \frac{e_{ij_1k_2}}{e_{ij_2k_2}} \tag{8}$$

A binary Y is now put to use. For any I we have $e_{+jk} = 1$, so when $i = 2$ we have $e_{2jk} = 1 - e_{1jk}$ for all j, k and from (8) we find that for all j_1, j_2, k_1, and k_2

$$\frac{e_{1j_1k_1}}{e_{1j_2k_1}} = \frac{e_{1j_1k_2}}{e_{1j_2k_2}} \tag{9}$$

$$\text{and} \quad \frac{(1 - e_{1j_1k_1})}{(1 - e_{1j_2k_1})} = \frac{(1 - e_{1j_1k_2})}{(1 - e_{1j_2k_2})} \tag{10}$$

Taking the value of $e_{1j_1k_1}$ from (9) and substituting it into (10) gives (after some algebraic manipulation) for all j_1, j_2, k_1, and k_2

$$(e_{1j_2k_2} - e_{1j_1k_2})(e_{1j_2k_1} - e_{1j_2k_2}) = 0 \tag{11}$$

In particular, consider the case $j_1 = j_*$, $j_2 = j_{**}$ and $k_2 = k_*$. It follows from (4) that $e_{1j_*k_*} \neq e_{1j_{**}k_*}$ and hence from (11) that $e_{1j_{**}k_*} = e_{1j_{**}k}$ for all k or, equivalently, for all k

$$\frac{e_{1j_{**}k_*}}{e_{1j_{**}k}} = 1$$

Because of (8), it must be that

$$\frac{e_{1jk_*}}{e_{1jk}} = 1 \quad \text{for all } k \text{ and for all } \quad j. \tag{12}$$

Choosing in one case $k = k'$ and in another $k = k''$ gives

$$1 = \frac{e_{1jk_*}}{e_{1jk'}} = \frac{e_{1jk_*}}{e_{1jk''}} \quad \text{for all } j$$

or equivalently $e_{1jk'} = e_{1jk''}$ for all j. Choosing $j = j'$ contradicts (5) and so proves the theorem. □

That the restriction to non-zero probabilities is immaterial can be seen as follows.

- If any marginal probability is zero, that category is removed and the proof follows as before.
- If a_{jk} is zero then because $X \perp\!\!\!\perp Z$, either a_{+k} or a_{j+} is zero and the appropriate category removed.
- If, say, $e_{1j_1k_1}$ is zero then the line immediately preceding (7) holds for this j_1 but (7) itself does not.

  In this case e_{1j_1k} is zero for all k which either contradicts (5) and the theorem is proved or j' cannot be this value of j_1. A similar argument shows that k_* cannot be this value of k_1. The proof proceeds with (7) exclusive of those values of i, j, and k for which $e_{ijk} = 0$.

That the theorem does not hold when $I > 2$ was shown by counter example in the text. The difficulty is that while (9) holds for all i when the first subscript on e is i rather than 1, (10) does not. Instead (10) becomes

$$\frac{(1 - \sum_{i=1}^{I-1} e_{ij_1k_1})}{(1 - \sum_{i=1}^{I-1} e_{ij_2k_1})} = \frac{(1 - \sum_{i=1}^{I-1} e_{ij_1k_2})}{(1 - \sum_{i=1}^{I-1} e_{ij_2k_2})}$$

which does not yield a factorization as in (11).

When $I = J = K = 2$, the notation of Table 1 can be used which considerably simplifies the proof. The proof of the theorem then becomes suitable as a mathematical problem for undergraduates.

Values for the counter-example of Figure 10

Using the notation introduced in the rubber wishbone theorem, Figure 10 was produced with the following values (varying the index j faster than k from left to right as in the leftmost eikosogram of Figure 10).

e_{1jk} :	$1/3$	$1/3$	$1/3$	$1/3$	$3/5$	$3/5$
a_{jk} :	$15/128$	$5/32$	$3/128$	$5/64$	$15/64$	$25/64$

References

Cherry, W.H. and R.W. Oldford (2002). "On the poverty of Venn diagrams for teaching probability: their history and replacement by Eikosograms." *(submitted for publication)*

Cox, D.R. and N. Wermuth (1996). *Multivariate Dependencies*. London, Chapman and Hall.

Dawid, A.P. (1979). "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society (B)*,

41, pp. 1-31.
Darroch, J.N., S.L. Lauritzen and T.P.. Speed (1980). “Markov Fields and Log-Linear Interaction Models for Contingency Tables”, *Annals of Statistics*, *8*, pp. 522-539.
Fienberg, S.E. (1977). *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA, MIT Press.
Lauritzen, S. (1996). *Graphical Models*. Oxford, Oxford University Press.
Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester UK, John Wiley & Sons.