COEFFICIENTS OF RELATION FOR PROBABILISTIC REASONING

Silvio Ursic Ursic Computing 810 Ziegler Rd. Madison. Wisconsin 53714

Definitions and notations with historical references are given for some numerical coefficients commonly used to quantify relations among collections of objects for the purpose of expressing approximate knowledge and probabilistic reasoning.

1. INTRODUCTION

Many coefficients that do not correspond to probabilities are being proposed and used for the purpose of considering uncertainty in artificial intelligence. As an example, consider the following assertion: "A car of brand 'A' develops transmission trouble three times more often than with other cars". Using P(X|Y) for the standard Bayesian conditional probability, one might have:

```
P( car has transmission problems | car brand A ) = .003, P( car has transmission problems | car not brand A ) = .001,
```

and we have our "three" as .003 / .001. If this ratio is one, we have statistical independence between the events "car brand A" and "car has transmission trouble". The coefficient "three" in the example above is a factor linking two conditional probabilities. Such factors are widely used in current expert system codes. They are known as "likelihood ratios", "certainty factors", "confidence factors", "evidential factors" (and others), as applied to "weights", "evidence", "beliefs", "rule strengths", "possibilities" (and others), with the purpose of "combine evidence", "maintain truth", "update beliefs", "propagate uncertainties" (and others).

2. SOME COEFFICIENTS OF RELATION

A small collection of numerical coefficients has been identified as forming the basic core of coefficients currently being used for uncertain reasoning. They are all old, simple and very useful. What follows does not introduce anything new, or original. Its purpose is notational and computational. The notation for conditional probabilities is universally accepted (and used). P(A|B) means P(AAB) / P(B) and not P(AAB) / P(A) or, say, P(-AAB) / P(B). Nevertheless, for other coefficients which express some relation between two statistical events distinct from the bayesian ratio, inconsistencies in actual usage remain strong. To be able to use these coefficients fluently, one cannot constantly have to worry if P(A,B) for one implementation is P(B,A), or P(A,B), or P(A,B) for another implementation. Hopefully, what follows should help develop a standard notation and nomenclature for them, which in turn should facilitate their usage.

This note also addresses the problem of making clearer the distinction

between what is being computed and the techniques used to compute it. Perhaps unintentionally, current efforts with approximate reasoning in artificial intelligence tend to blend the two subjects together. The net result of this state of affairs may have a desirable effect from a marketing point of view. In the resulting confusion, it is always possible to claim that "our package" computes something totally different (and much better) than "their package", with no possibility of being contradicted. The discussion following the presentation of [Wise 86] at the 1986 Uncertainty in Artificial Intelligence Workshop was quite representative of this, and illuminating. The heated debate that resulted is a direct consequence of this blending of problems with methods of solution.

To reiterate the point, what needs to be computed to simulate probabilistic aspects of human intelligence is a psychological or perhaps a philosophical problem. Coefficients of relation have been found useful for this purpose. How to carry out the computations specified by these psychological and philosophical considerations is a problem of a rather distinct nature. The specific numerical and combinatorial methods employed to solve some proposed system of numerical equations will no doubt influence computing times, memory utilization, problems with numerical stability, precision, etc. Ultimately, the available methods of solution determine what can and what cannot be done. The methods of solution should not, however, subreptitiously change what is supposedly being computed.

The definitions and notation for the basic core of coefficients of relation for reasoning with uncertainty are:

```
_______
  PROBABILITY
  P(A)
                        Classical point probability
  P(A|B) = P(ABB) / P(B)
                        Conditional probability (Bayesian)
  INDEPENDENCE (Quetelet)
  Q(A|B) = O(A|B) / O(A|-B)
                        Conditional Odds Ratio
  Q(A:B) = P(A|B) / P(A|-B)
                        Conditional Probability Ratio
  EXCHANGEABILITY (de Finetti)
  F(A|B) = O(A|B) / O(B|A)
                        Conditional Odds Ratio
  F(A:B) = P(A|B) / P(B|A)
                        Conditional Probability Ratio
------
```

We use here "-" for a logical NOT and "&" for a logical AND. The survey [Goodman & Kruskal 1954, 1959] will give the reader a taste of the variety of functions on the four quantities P(A&B), P(A&-B), P(-A&B) and P(-A&-B) that have been found useful for some purpose. The paper [Fienberg & Gilbert 1970] provides interesting geometric insights on independence. Also see [Good 1965, 1985] for additional stimulating information.

3. RANGES

Mainly for psychological reasons, it has repeatedly been found necessary to provide alternate ranges to the zero-one and zero-infinite ranges of these coefficients. In particular, a "minus one to plus one" range has been found many times very appealing. The most commonly used ranges and conversions are:

```
probability range (P-type): zero to one;
odds range (O-type): zero to infinity;
symmetric range (S-type): minus one to plus one.
```

CONVERSIONS AMONG RANGES

	FROM:	P-type	0-type	S-type
TO:	∫ = =			
P-type	ĺ	P	0 / (1+0)	(S+1) / 2
0-type	ĺ	P / (1-P)	0	(1+S) / (1-S)
S-type	ĺ	2*P - 1	(0-1) / (0+1)	S
	Ĺ==			j

With these ranges, each coefficient can have two additional aliases. For probabilities, however, the symmetric range has been found inconvenient. Similarly, for the independence and exchangeability coefficients, the probability range does not seem useful. The notations for the useful aliases are:

<u>MAIN</u>				ALTERNATE		
P(A)	P-type	[0, 1]	O(A)	0-type	[0, infinite]	
P(A B)	P-type	[0, 1]	O(A B)	0-type	[0, infinite]	
Q(A B)	0-type	[O, infinite]	QS(A B)	S-type	[-1, +1]	
Q(A:B)	0-type	[O, infinite]	QS(A:B)	S-type	[-1, +1]	
F(A B)	O_type	[O, infinite]	FS(A B)	S-type	[-1, +1]	
F(A:B)	0-type	[O, infinite]	FS(A:B)	S-type	[-1, +1]	

In general, if a / b is a coefficient of odds type then a / (a+b) is of probability type and (a-b) / (a+b) is of symmetric type. From a computational perspective, alternate ranges are not needed. However, expressing degrees of independence and exchangeability with a coefficient between minus one and one, with a zero indicating independence or exchangeability, seems to sometimes tickle our brain the right way. In many situations we also find that odds are favored over probabilities (see, for example, betting). Perhaps this is how some coefficients of relation are biologically stored, and hence further conversions to the odds scale or to the probability scale involves additional "thinking", which we prefer to avoid. One may classify the situation as one of wanting biologically pleasing units of measure at our disposal.

4. PRODUCT PARTITION DEFINITIONS

Let:	В -В	
P(A & B) = x,	++	+
P(A & -B) = y,	A x y	١
P(-A & B) = z,	++	+
P(-A & -B) = w.	-A Z W	١
	++	+

We then have:

$\frac{PROBABILITY}{P(A) = 1 - P(-A) = x + y} \\ P(B) = 1 - P(-B) = x + z$ $\frac{ODDS}{O(A) = 1 / O(-A) = (x + y) / (z + w)} \\ O(B) = 1 / O(-B) = (x + z) / (y + w)$

```
CONDITIONAL PROBABILITY
```

```
P(A | B) = 1 - P(-A | B) = x / (x + z)

P(A | -B) = 1 - P(-A | -B) = y / (y + w)

P(B | A) = 1 - P(-B | A) = x / (x + y)

P(B | -A) = 1 - P(-B | -A) = z / (z + w)
```

CONDITIONAL ODDS

```
O(A| B) = 1 / O(-A| B) = x / z
O(A|-B) = 1 / O(-A|-B) = y / w
O(B| A) = 1 / O(-B| A) = x / y
O(B|-A) = 1 / O(-B|-A) = z / w
```

QUETELET PROBABILITIES RATIO

QUETELET ODDS RATIO

```
Q(A|B) = Q(B|A) = Q(-A|-B) = Q(-B|-A) = 1/Q(A|-B) = 1/Q(B|-A) = 1/Q(-A|B) = 1/Q(-B|A) = xw / yz
```

de FINETTI PROBABILITIES RATIO

```
F(A: B) = 1 / F(B: A) = (x + y) / (x + z)

F(A:-B) = 1 / F(-B: A) = (x + y) / (y + w)

F(-A: B) = 1 / F(B:-A) = (z + w) / (x + z)

F(-A:-B) = 1 / F(-B:-A) = (z + w) / (y + w)
```

de FINETTI ODDS RATIO

```
F(A|B) = F(-B|-A) = 1 / F(B|A) = 1 / F(-A|-B) = y / z

F(-A|B) = F(B|-A) = 1 / F(B|-A) = 1 / F(-A|B) = x / w
```

Notice that we have four distinct coefficients of type Q(A:B), F(A:B) and P(A|B), two of type F(A|B) and only one of type Q(A|B). From a computational perspective, Q(A|B) and F(A|B) are both cheaper than Q(A:B) and P(A:B). They are the preferred form for independence and exchangeability. The independence of A and B can be declared both with Q(A|B) = 1 and Q(A:B) = 1. Similarly, the exchangeability of A and B can be set with P(A|B) = 1 or P(A:B) = 1. For values near one, the pair P(A|B) and P(A:B) and also the pair P(A|B) and P(A:B) convey almost the same information.

5. CONDITIONAL PROBABILITY

Thanks to [Bayes 1763], the notation $P(A \mid B)$ is a universal standard for conditional probability.

6. INDEPENDENCE

The papers [YULE 1900, 1903, 1912] introduced the "Q" notation for these two independence coefficients. Yule choose "Q" in honor of Quetelet. The book translation from the French [Quetelet 1849] is one of the few references given by Boole as having had an impact on his writings in probability. The deciphering of the very verbose exposition in Quetelet book might have been Boole's motivation to develop the compact notation that we today know as "Boolean algebra". with the heavy emphasis on probabilities he gave to his

work. In connection with independence of statistical events, the paper [Wilbraham 1854] can be considered a classic in misinterpretation of its In general, one does not know if two given events are applicability. independent. By default their relation is unknown. This limited role assumed by statistical independence has been a stumbling block to the applicability of existing statistical results for artificial intelligence and uncertain A second reason for the unwarranted distaste shown by some reasoning. researchers in AI towards long established ideas and results in probability can be traced to plain and simple "algorithmic draught". The computational problems that arise when we wish to implement ideas and principles exemplified with three or four events to three or four thousand events necessitate the development of entirely new algorithmic approaches.

7. EXCHANGEABILITY

Exchangeability introduces many properties that mirror the properties of independence. For example, with N exchangeable events we have N degrees of freedom among the 2^N product partition events, exactly as with independent events. Consult [de Finetti 1937, 1969], [Chrisma 1971, 1982], [Diaconis 82], [Galambos 82] to get acquainted with results and a growing bibliography on exchangeability. Many statistical events in approximate reasoning are more precisely modeled by assuming exchangeability than independence. The constraints introduced by exchangeability blend easily with the polyhedral constraints arising from the fact that the statistical events under consideration are at the onset defined with arbitrary boolean functions, which are also polyhedral constraints. Consult [Ursic 1987] for additional information on the constraints on the probabilities of the product partition arising from boolean constraints.

8. CONCLUDING REMARKS

It seems inevitable that many other coefficients, besides the ones reported here, will become necessary to quantify finer and finer properties for relations among events. For example, chi-square for two events is also an independence coefficient. However, the coefficients Q(A|B), Q(A:B), P(A|B) and P(A:B) stand out for their simplicity and generality.

In using these coefficients for approximate reasoning one tacitly assumes the validity of probabilistic ideas as a framework for reasoning with uncertainty. Insights on this can be found in [Cheeseman 1985]. A further consequence of the probabilistic approach is the demise of the concept of a logical deduction. All what we truly have is a system of linear and nonlinear equations, some derived from coefficients of relation and some derived from boolean definitions of events. As an example of this, consult [Dempster 1967], [Quinlan 1983, 1985]. The sought answers, in the form of probabilities of some events and of coefficients of relation for some collections of events, are simply obtained by solving appropriate systems of algebraic equations. If possible, we solve them by analytical means. If analytical methods fail, as has been the case with many probabilistic problems arising from artificial intelligence, we solve them with numerical methods. If exact solutions cannot be obtained in reasonable amounts of time, we use approximate methods.

Describing Gaussian elimination or the simplex method as procedures performed by an inference net, or by an algorithm performing a deduction, or by a forward or backward chain of inferences does not seem to describe the methods of solution that can be actually employed. Efficient methods of solution for

these systems of equations may have very little relation with what they mean to the end user. As a point fo reference, consider the fast Fourier transform algorithm and its relation to some intuitive meaning of harmonic analysis. The point of view taken here and in [Ursic 87] is that most (if not all) of what is currently being proposed as reasoning with uncertainty can be interpreted as consisting of the problem of solving systems of linear and nonlinear inequalities and equations with the unknowns being some sought collection of coefficients of relation. Advances will therefore come from two sources. First, the equations to be solved must be clearly and precisely stated. is not an easy task, especially with respect to the constraints arising from logical conditions. Second, we systematically develop specialized numerical methods for the solution of the particular systems of equations so obtained. As a consequence, with the help of standardized test problems, available methods of solution, trade-offs between precision and computing times, specialized sub-problems, etc., can be analyzed and compared.

From this perspective, most (if not all) of the current efforts in approximate reasoning can be interpreted as being ad hock methods of solution (neither very efficient nor very precise) for the very special systems of equations and inequalities arising from the logical and statistical constraints defining the problem at hand. Trying to mimic what we perceive as being the solution methods employed by biological computers for the problem may be counterproductive. Biological systems have such severe limitations in energy consumption and energy density that the solutions they developed to the problems we are facing may not be suitable to the tools at our disposal. One should consider that we do not build airplanes with flapping wings.

REFERENCES

- Bayes T., An essay towards solving a problem in the doctrine of chances. Phil. Trans. Roy. Soc. 53 (1763) 370-418. Reprint in: Biometrika, 45 (1958) 296-315.
- Cheeseman P., In defense of probability. 9th International Joint Conference on Artificial Intelligence, Los Angeles, 1985.
- Crisma L., Alcune valutazioni quantitative interessanti la proseguibilita di processi aleatori scambiabili. Rendiconti Istituto di Matematica, Università di Trieste, 3 (1971) 1-29.
- Crisma L., Quantitative analysis of exchangeability in alternative processes. Exhangeability in Probability and Statistics, Koch G., Spizzichino F., eds. North Holland Company, 1982.
- Dempster A.P., Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statist., 38 (1967) 325-339.
- Diaconis P., Finite forms of de Finetti's theorem of exchangeability. Synthese 36 (1977) 271-281.
- Diaconis P., Zabell S.L., Updating subjective probability. Journal of the American Statistical Association, 77 (1982) 822-830.
- Duda R.O., Hart P.E., Nilsson N.J., Subjective Bayesian methods for rule-based inference systems. AFIPS Conference Proceedings, National Computer Conference, 1976, pp 1075-1082.
- Fienberg S.E., Gilbert J.P., The geometry of a two by two contingency table. Journal of the American Statistical Association, 65 (1970) 694-701.
- de Finetti B., La prevision: ses lois logiques, ses sources subjectives.

 Annales de l'Institut Henri Poincare, 7 (1937) 1-68. Translated as:
 Foresight: its logical laws, its subjective sources. In: Studies in subjective Probability, Kyburg Jr.H.E., Smokler H.E. eds. John Wiley & Sons, Inc., New York, 1964.

- de Finetti B., Sulla proseguibilita di processi aleatori scambiabili. Rendiconti Istituto di Matematica, Università di Trieste, 6 (1969) 53-67.
- Galambos J., The role of exchangeability in the theory of order statistics. Exchangeability in Probability and Statistics, Koch G., Spizzichino F., eds. North Holland Publishing Company, 1982.
- Good I.J., The estimation of probabilities. Research Monograph No. 30, The M.I.T. Press, Cambridge, Massachusetts, 1965.
- Good I.J., Mathematically natural generalizations of some measures of association and dependence for contingency tables. J. Statist. Comput. Simul., 22 (1985) 93-97.
- Goodman L.A., Kruskal W.H., Measures of association for cross classifications. Journal of the American Statistical Association Journal, 49 (1954) 732-764.
- Goodman L.A., Kruskal W.H., Measures of association for cross classifications. II: Further discussion and references. Journal of the American Statistical Association Journal, 54 (1959) 123-163.
- Koch G., Spizzichino F., Exchangeability in Probability and Statistics, Proceedings of the international conference, Rome, April 1981. (1982) North-Holland Publishing Company.
- Popper K., Miller D., A proof of the impossibility of inductive probability.
 Nature 302 (1983) 687-688. Comments in: Nature, Matters Arising, 310 (1984) 433-434.
- Pratt J.W., Raiffa H., Schlaifer R., The foundations of decision under uncertainty: an elementary exposition. American Statistical Association Journal, 59 (1964) 353-375.
- Quetelet M. A., Sur la possibilité de mesures l'influence des causes qui modifient les éléments sociaux, (1832) pamphlet. (Cited by Yule (1900) to justify the "Q" notation.)
- Quetelet A., Letters addressed to H. R. H. the Grand Duke of Saxe Coburg and Gotha, on the theory of probabilities: as applied to the moral and political sciences, translated from the french by O.G. Downes, London, C.&E. Layton 1849.
- Quinlan J.R., Inferno: a cautious approach to uncertain inference. The Computer Journal, 26 (1983) 255-269.
- Quinlan J.R., Internal consistency in plausible reasoning systems. New Generation Computing, 3 (1985) 157-180.
- Ursic S., A linear characterization of NP-complete problems. Seventh International Conference on Automated Deduction, Springer Verlag Lecture Notes in Computer Science 170 (1984) 80-100.
- Ursic S., Generalizing fuzzy logic probabilistic inferences. Uncertainty in Artificial Intelligence, North Holland, 1987.
- Venn J., Symbolic Logic. Macmillan and Co., London, 1881.
- Wilbraham H., On the theory of chances developed in Professor Boole's "Laws of Thought". Philosophical Magazine and Journal of Science, Supplement to vol. VII, Fourth Series, 48 (1854) 465-476.
- Wise B.P., Experimentally comparing uncertain inference systems to probability. Uncertainty in Artificial Intelligence, North Holland, 1987.
- Yule G.U., On the association of attributes in statistics: with illustrations from the material from the Childhood Society, &c., Phil. Trans. Roy. Soc., Ser. A, 194 (1900) 257-319. (Reprinted in "Statistical papers of George Udny Yule", edited by M.G. Kendall and A. Stuart, Hafner Publishing Company, New York 1971.)
- Yule G.U., Notes on the theory of association of attributes in statistics, Biometrika 2, (1903) 121-134.
- Yule G.U., On the methods of measuring association between two attributes. Journal of the Royal Statistical Society, 75 (1912) 579-642, 643-52.