AXIOMATIC APPROACHES TO FORMULAS FOR COMBINING LIKELIHOODS OR EVIDENCE

NORMAN L. JOHNSON
University of North Carolina
at Chapel Hill

SAMUEL KOTZ
University of Maryland
College Park

ABSTRACT

An elementary axiomatic approach along the lines of axiomatization of probability functions is developed. Some essential differences are noted.

Key Words: Axiomatization; Combination of Evidence, Support.

1. INTRODUCTION

Combinations of 'evidence', 'expert opinion', 'support', 'likelihoods' or just 'data' in general, in a fundamental problem of statistical analysis, with an implied emphasis on condensation and summarization. Although the problem has a venerable history, dating back at least as far as T. Bayes' investigations, and has, indeed, received considerable attention in the statistical literature of the last two decades (e.g. McConway (1981), Good (1983,1985), Genest and Zidek (1986)), this theme has attracted only superficial attention in most textbooks. Moreover, much current research in the field seems to be devoted, primarily, to construction of rules for combination ('pooling') of opinions in specific cases (see, e.g. Stone (1961), and later, more detailed discussions in DeGroot (1974) and Kadane et al. (1980)). To the best of our knowledge, there have been few serious attempts at axiomatization of combination of evidence analogous to Kolmogorov's axiomatization of probability or Shannon's axiomatization of measures of information.

It was against this background that our attention was caught by an article by V.L. Stefanyuk (1987) devoted to a description of expert systems. He presented an axiomatic approach to certain formulas for combinations of evidence (or likelihoods). In particular he derived an analog of the classical addition formula for probabilities. This derivation, however, was based on quite sweeping assumptions about the function representing support ('degree of confidence' is Stefanyuk's term) for a hypothesis.

In this paper we (i) derive Stefanyuk's formula in a simpler fashion with less restrictive assumptions and (ii) investigate the effect of further relaxation of a key assumption.

Related arguments will be found in Shortliffe and Buchanan (1975) and Shortliffe (1976, Chapter 4) in connection with combination of expert opinions in medical diagnosis. See also the recent article by Rennelles and Shortliffe (1987).

The elementary deviations and analysis in this note may be of use in courses on probability theory for undergraduates and first year graduate students, and may trigger further research in this field.

2. STEFANYUK'S AXIOMS

Stefanyuk (1987) presented the following axioms for a function $x(\alpha,\beta)$ representing the combination of two supports α,β for a hypotheses:

(a) $0 \le \alpha \le 1$; $0 \le \beta \le 1$; $0 \le x(\alpha,\beta) \le 1$.

(b) $x(\alpha,0) = \alpha$

(c) $x(\alpha,\beta)$ is a symmetric function of α and β.

(d) $x(1,1) = 1$.

(e) $x(\alpha,\beta)$ can be expanded as a power series in α and β.

(f) subject to (a)–(e), $x(\alpha,\beta)$ has its maximum possible value.

Axiom (a) and (b) establish a scale of 0 to 1 for 'support'. (Note that this approach does not allow for negative support, though Shortliffe (1976) does allow for this – see Section 6.)

Axiom (c) is natural; it supposes that the order in which the support is received is immaterial. Axiom (d) just says that the combination of two evidences, each giving full support separately, also give full support when combined.

The remaining axioms are less self–evident. Axiom (e) is, perhaps, not unreasonable; though Stefanyuk (and ourselves) restrict it further in order to obtain explicit formulas for $x(\alpha,\beta)$. Axiom (f) seems the least natural, though it is in the spirit of axiomatizations along information–theoretic lines. In fact in our derivation of Stefanyuk's formula, using his restricted form of (e), we do not need to use (f). It is, however, needed when the restrictions on (e) are relaxed slightly.

An equally appealing axiom would be (f)' $x(\alpha,\beta)$ is a nondecreasing function of α and β separately. (Note that if (f) were to be replaced by (f)', then axiom (d) would not be necessary, since from (b) we would have $x(1,0) = 1$, and from (f)' and (a), $1 \leq x(1,1) \leq 1$.

3. STEFANYUK'S FORMULA

Stefanyuk restricts the expansion under axiom (e) to terms of 2nd and lower order, and then derives the formula

$$x(\alpha,\beta) = \alpha + \beta - \alpha\beta. \tag{1}$$

This has a formal analogy to the well–known formula,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

for the probabilities relating to two independent events A,B.

Axiom (e) states that

$$x(\alpha,\beta) = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} a_{rs} \alpha^r \beta^s .$$

With Stefanyuk's restriction, $a_{rs} = 0$ for $r+s > 2$.

We first note that, generally

(i) (c) implies $a_{rs} = a_{sr}$.

(ii) (b) implies $\sum_{r=0}^{\infty} a_{r0}\, \alpha^r \equiv \alpha$ for $0 \leq \alpha \leq 1$, whence

$$a_{10}(-a_{01}) = 1; \quad a_{r0}\ (=a_{0r}) = 0 \quad \text{for } r \neq 1.$$

(iii) (d) implies

$$\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} a_{rs} = 1 = \sum_{r=0}^{\infty} a_{rr} + 2 \sum_{1\leq r<s}^{\infty}\sum^{\infty} a_{rs} + 2 \sum_{s=1}^{\infty} a_{0s} \qquad \text{(cf. (i))}$$

$$= \sum_{r=1}^{\infty} a_{rr} + 2 \sum_{1\leq r<s}^{\infty}\sum^{\infty} a_{rs} + 2 \qquad \text{(cf. (ii))}$$

whence

$$\sum_{r=1}^{\infty} a_{rr} + 2 \sum_{1\leq r<s}^{\infty}\sum^{\infty} a_{rs} = -1 .$$

If $a_{rs} = 0$ for $r+s > 2$, then (iii) gives $a_{11} = -1$, and so, noting (from (ii)) that $a_{10} = a_{01} = 1$, we have

$$x(\alpha,\beta) = \alpha + \beta - \alpha\beta$$

which is Stefanyuk's formula.

Note that condition (f) is not used in this derivation.

4. CASE r+s < 3

If the restriction on (e) is relaxed to allow $a_{rs} \neq 0$ if $r+s \leq 3$ then from (i)–(iii):

$$x(\alpha,\beta) = \alpha+\beta - \alpha\beta - a_{12}\, \alpha\beta(2-\alpha-\beta)$$

Since $2-\alpha-\beta > 0$ if $\alpha \neq \beta$, condition (f) requires that a_{12} has the least possible value.

Now $x(1,\beta) = 1-a_{12}\,\beta(1-\beta) \leq 1$.

Hence $a_{12} \geq 0$, and the least possible value of a_{12} is 0.

So $x(\alpha,\beta) = \alpha+\beta-\alpha\beta$.

This is the same formula as for $r+s \leq 2$; however, we will now show that a different result is obtained if $r+s \leq 4$.

5. CASE $r+s \leq 4$

If we allow $a_{rs} \neq 0$ for $r+s \leq 4$, then from (i)–(iii):

$$x(\alpha,\beta) = \alpha+\beta-\alpha\beta - a_{12}\,\alpha\beta(2-\alpha-\beta) - a_{13}\,\alpha\beta(2-\alpha^2-\beta^2) - a_{22}\alpha\beta(1-\alpha\beta);$$

$$x(1,\beta) = 1-(a_{12}+a_{22})\beta(1-\beta) - a_{13}\beta(1-\beta^2)$$

In order to satisfy (f) we will make $x(1,\beta) \equiv 1$ for $0 \leq \beta \leq 1$. Equating coefficients of powers of β to zero, we obtain

$$a_{12} + a_{22} + a_{13} = a_{12} + a_{22} = a_{13} = 0$$

whence

$$x(\alpha,\beta) = \alpha+\beta-\alpha\beta - a_{12}\alpha\beta(1-\alpha-\beta+\alpha\beta)$$

$$= \alpha+\beta-\alpha\beta - a_{12}\,\alpha\beta(1-\alpha)(1-\beta)$$

To satisfy (f) we require a_{12} to have its least possible value.

Since $x(1,\beta) \equiv 1$ for all β and $x(\alpha,\beta) \leq 1$ for all α,β we must have

$$\left.\frac{\partial x(\alpha,\beta)}{\partial\alpha}\right|_{\alpha=1} \geq 0 \quad \text{for all } \beta.$$

i.e. $(1-\beta)(1+a_{12}\beta) \geq 0$

whence

$$a_{12}\,\beta \geq -1$$

i.e. $a_{12} \geq -1/\beta$ for $0 \leq \beta \leq 1$.

The least possible value of a_{12} is therefore –1, giving

$$\begin{aligned} x(\alpha,\beta) &= \alpha+\beta-\alpha\beta+\alpha\beta(1-\alpha)(1-\beta) \\ &= [= 1-(1-\alpha)(1-\beta)(1-\alpha\beta)] \end{aligned} \qquad (2)$$

This differs from Stefanyuk's formula by the additional term $\alpha\beta(1-\alpha)(1-\beta)$.

6. THE POSSIBILITY OF ALLOWING FOR NEGATIVE SUPPORT

Although the axioms (a)–(f) do not require $x(\alpha,\beta)$ to be a nondecreasing function of α and β separately, both formulas (1) and (2) do have this property. It is, therefore, implicit in this approach that any additional evidence cannot decrease the total support – that is, the possibility of adverse evidence is excluded. We now consider some points arising in trying to allow for adverse evidence.

We present the following axioms, tentatively, as a basis for allowing for negative support ("evidence against")

(a)* $\quad -1 \leq \alpha \leq 1;\ -1 \leq \beta \leq 1;\ -1 \leq x(\alpha,\beta) \leq 1.$

(g) $\quad x(-\alpha, -\beta) = -x(\alpha,\beta)$

together with (b), (c) and (d), and possibly (f).

Note that from (g)

$$x(-\alpha,\alpha) = -x(\alpha,-\alpha)$$

but from (d) $\quad x(-\alpha,\alpha) = x(\alpha,-\alpha)$

so $x(\alpha,-\alpha) = 0$ which is natural, as it reflects a balance of evidence pro and con.

The simple approach of converting formula (1) to allow for a measure α^* of support having range –1 to +1, by the transformation

$$\tfrac{1}{2}(\alpha^*+1) = \alpha, \text{ or equivalently } \alpha^* = 2\alpha-1$$

would not satisfy (g).

In fact, from (1) we would obtain

$$\tfrac{1}{2}\{x^*(\alpha^*,\beta^*) - 1\} = \tfrac{1}{2}(\alpha^*+1) + \tfrac{1}{2}(\beta^*+1) - \tfrac{1}{4}(\alpha^*+1)(\beta^*+1)$$

whence

$$x^*(\alpha^*,\beta^*) = \tfrac{1}{2}(1+\alpha^*+\beta^*-\alpha^*\beta^*),$$

implying

$$x^*(\alpha^*,-\alpha^*) = \tfrac{1}{2}(1+\alpha^{*2}) \neq 0$$

and $\quad x^*(-\alpha^*,-\beta^*) = \tfrac{1}{2}(1-\alpha^*-\beta^*+\alpha^*\beta^*) \neq -x^*(\alpha^*,\beta^*).$

Note, also, that the natural conditions

$$x^*(1,\beta^*) = 1 \quad \text{for } \beta^* > 0; \quad x^*(1,\beta^*) < 1 \quad \text{for } \beta^* < 0$$

cannot be satisfied by any polynomial representation. (If the polynomial $x^*(1,\beta^*)$ is constant for $0 < \beta^* \leq 1$, it must be constant for all β^*, including $\beta^* < 0$.)

It does seem desirable to allow for the possibility of support decreasing as a result of additional information. There are, however, circumstances – for example, when support is measured by precision, or reciprocal variance of an estimator – when additional (independent) observation cannot decrease support, so that Stefanyuk's approach can be justified. Another example is in multiple linear regression analysis, wherein the addition of a further predicting (control) variable cannot reduce the residual sum of squares (of differences between observed and fitted regression values).

REFERENCES

DeGroot, M.H.(1974) Reaching a consensus, J. Amer. Statist. Ass., 69, 118–121.

Genest, C. and Zidek, J.V. (1986) Combining probability distributions, Statist. Sci, 1, 114–135.

Good, I.J. (1983) Good Thinking: The Foundations of Probability and its Applictions, Minnesota University Press, Minneapolis, Minnesota.

Good, I.J. (1985) Some statistical applications of Poisson's work, Statist. Sci., 1, 157–180.

Kadane, J.B., Dickey, J.H., Winkler, R.L., Smith, W.S. and Peters, S.C. (1980) Interactive elicitation of opinion for normal linear model, J. Amer. Statist. Ass., 75, 845–854.

McConway, K.J. (1981) Marginalization of linear opinion pools, J. Amer. Statist. Ass., 76, 410–414.

Renelles, G.D. and Shortliffe, E.H. (1987) Advanced computing for medicine, Sci. American, 257 (4), 154–161.

Shortliffe, E.H. (1976) Computer Based Medical Consultations. MYCIN, American Elsevier, New York.

Shortliffe, E.H. and Buchanan, B.G. (1975) A model of inexact reasoning in medicine, Math. Bio Sci., 23, 351–379.

Stefanyuk, V.L. (1987) Some aspects of theory of expert systems, Soviet J. Comp. Syst. Sci., 25(2), (translation of original Russian article in Tekh. Kibernet., No. 2, 85–91 (1987)).

Stone, H. (1961) The opinion pool, Ann. Math. Statist., 32, 1339–1342.