Factoring bivariate distributions

Frank Lad
University of Canterbury*

Abstract

Every probability distribution over two events is shown to have a representation for which the probability mass function is factored into the marginal mass functions via a parametric equation. We study the coherent parameter space and the restrictions on probability distributions induced by any parameter specification. All results are illuminated by graphic displays.

Key Words: Factoring probabilities, exchangeability, extendibility, independence, coherence, probability bounds.

1 Motivation

Common practice in statistical modelling often presumes the stochastic independence of events for the mathematical convenience of the factorisation of $P(AB)$ into $P(A)P(B)$ rather than for realistic merit. This article shows that every coherent distribution over the partition generated by A and B has a factorised representation in terms of the marginal probabilities via the parametric equation $P(AB) = P(A)^{\alpha}P(B)^{\beta}$. We study the parameter space of allowable pairs of (α, β) and the restrictions on $(P(A), P(B), P(AB))$ induced by each specification of (α, β).

2 Results

For any two events A and B, consider the restrictions on probability assertions P(A), P(B) and P(AB) induced by the relation

$$P(AB) \ = \ P(A)^{\alpha} \ P(B)^{\beta} \ ,$$

for some real parameters α and β . When $\alpha = \beta = 1$, this is the relation specifying an assertion of independence regarding A and B, which allows the pair $(P(A), P(B))$ to reside anywhere in the unit-square. The numbered statements that follow indicate provable properties of the parameter space of coherent (α, β) specifications and of the probability assertions $(P(A), P(B))$ they support.

1. The specification that $P(AB) = P(A)^{\alpha}P(B)^{\beta}$ implies the following coherency conditions on probabilities over the partition generated by A and B:

$$\begin{array}{rlcll}
i. & p_{00} \equiv & P(\tilde{A}\tilde{B}) & = 1 - P(A) - P(B) + P(A)^{\alpha}P(B)^{\beta} & \\
ii. & p_{01} \equiv & P(\tilde{A}B) & = P(B) - P(A)^{\alpha}P(B)^{\beta} & \\
iii. & p_{10} \equiv & P(A\tilde{B}) & = P(A) - P(A)^{\alpha}P(B)^{\beta} \ , & \text{and} \\
iv. & p_{11} \equiv & P(AB) & = P(A)^{\alpha}P(B)^{\beta}. &
\end{array}$$

2. The parameter space for (α, β) supporting coherent specifications of $(p_{00}, p_{01}, p_{10}, p_{11})$ within the unit simplex is a symmetric subspace of $\Re^2$ with five symmetrically paired parts:

I. $0 < \alpha < 1, \quad \beta \geq \alpha, \quad \text{and} \quad \alpha + \beta \geq 1$;

*Address for correspondence: Frank Lad, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand. Email: F.Lad@math.canterbury.ac.nz

II. $\beta \geq \alpha \geq 1$;
III. $\alpha \in (0,1)$, and $\beta > 1$;
IV. $\alpha < 0$, $\beta > 1$, and $\alpha + \beta \geq 1$; and
V. $\alpha < 0$, $\beta > 1$, and $\alpha + \beta < 1$.

The five parts of this space are displayed in Figure 1-9, labeled appropriately by Roman numerals **I** through **V**, along with their symmetric reflections which are labeled similarly with primes (') attached, viz. **I'** through **V'**.

The symmetry of the parameter space of (α, β) pairs arises naturally from the fact that the designation of the two events by letters A and B is arbitrary. We can confine our attention to the half-space in which $\beta \geq \alpha$.

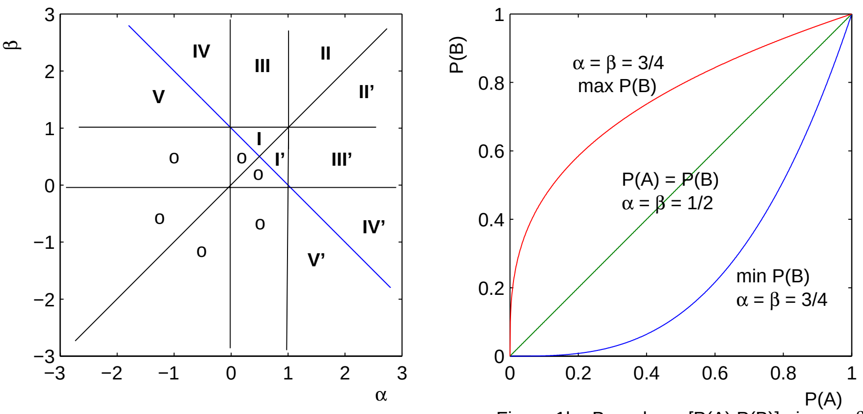

Figure 1a. The parameter space for (α,β) is composed of five regions, labeled **I** through **V**, along with their symmetric reflections about the line β = α. Regions marked with an "o" contain incoherent specifications of (α, β) pairs.

Figure 1b. Bounds on [P(A),P(B)] given α, β exemplified for α = β = 3/4. In the case of α = β = 1/2 it is necessary that P(A) = P(B). When α = β = 1, (P(A),P(B)) may lie anywhere within the unit−square.

Figure 1: The figure at left, 1a, portrays the parameter space of coherent (α, β) pairs described in statement **2**. The figure at right, 1b, partrays bounds on $(P(A), P(B))$ pairs that cohere with the specification of $\alpha = \beta = 3/4$ and $\alpha = \beta = 1/2$ as described in statement **3.I**.

3. Coherency conditions on probabilities specified in **1** imply distinct bounds on pairs of $(P(A), P(B))$ that are representable by (α, β) pairs in each of the five regions specified in **2**:

I. $P(A)^{\frac{\alpha}{1-\beta}} \leq P(B) \leq P(A)^{\frac{1-\alpha}{\beta}}$. This condition derives from the requirements that $P(A)^{\alpha}P(B)^{\beta} \leq P(A)$ and $P(A)^{\alpha}P(B)^{\beta} \leq P(B)$. The exponents on $P(A)$ in the boundary inequality must satisfy $\frac{\alpha}{1-\beta} \geq \frac{1-\alpha}{\beta}$, since $P(A) \in [0,1]$, as is $P(B)$. Equivalently, $\alpha + \beta \geq 1$.

II. $P(A)^{\alpha}P(B)^{\beta} \leq P(A) + P(B) \leq P(A)^{\alpha}P(B)^{\beta} + 1$. This is an implicit functional inequality on $(P(A), P(B))$ deriving from the condition that $p_{00} \equiv P(\tilde{A}\tilde{B}) \in [0,1]$.

III. $0 \leq P(B) \leq P(A)^{\frac{1-\alpha}{\beta}}$. This derives solely from the requirement that $P(A)^{\alpha}P(B)^{\beta} \leq P(A)$. Notice that the exponent on $P(A)$ in the bounding inequality for $P(B)$ is $\frac{1-\alpha}{\beta}$ which is always less than 1 in this region. Thus, the bounding function here is concave downward but is always exceeding $P(A)$.

IV. Here again, $0 \leq P(B) \leq P(A)^{\frac{1-\alpha}{\beta}}$. Coherency conditions would imply both that $P(B) \leq P(A)^{\frac{1-\alpha}{\beta}}$ and that $P(B) \leq P(A)^{\frac{\alpha}{1-\beta}}$. However these upper bounds on $P(B)$

are ordered by $P(A)^{\frac{1-\alpha}{\beta}} \leq P(A)^{\frac{\alpha}{1-\beta}}$, deriving from the fact that $\alpha + \beta \geq 1$. As in region **III**, the bounding function on $P(B)$ is concave downward, always exceeding $P(A)$.

V. $0 \leq P(B) \leq P(A)^{\frac{\alpha}{1-\beta}}$. Here the bounds on $(P(A), P(B))$ are derived exactly as in region **IV**. However, in this region, $P(A)^{\frac{1-\alpha}{\beta}} > P(A)^{\frac{\alpha}{1-\beta}}$, for here $\alpha + \beta < 1$. The bounding function on $P(B)$ is now a convex function of $P(A)$, lying always beneath the 45° line of $P(B) = P(A)$.

The parts of $\Re^2$ that are excluded from the parameter space for (α, β) are eliminated because they allow only specifications of $(P(A), P(B))$ that are incoherent. It is obvious that when α and β are both negative, $P(A)^\alpha P(B)^\beta$ would exceed 1. Secondly, $P(A)^\alpha P(B)^\beta$ would always exceed $P(B)$ when $\alpha < 0$ and $\beta \in (0, 1)$. Finally, when $\beta \geq \alpha > 0$ and $\alpha + \beta < 1$, coherency would require the inequality $P(A)^{\frac{\alpha}{1-\beta}} < P(B) < P(A)^{\frac{1-\alpha}{\beta}}$, as in region **I**. However under these parameter configurations this lower bound would have to exceed the upper bound.

Figure 10 can be used to understand the type of bounds on $(P(A), P(B))$ pairs when (α, β) lies regions I, III, IV or V. In region **I**, $(P(A), P(B))$ would be bounded within a region such as typified by the $minP(B)$ and the $maxP(B)$ functions shown in the Figure. (If $\alpha \neq \beta$ then the min and max functions would not be symmetric, but they would always straddle the line $P(B) = P(A)$.) In regions **III** and **IV** it is only required that $(P(A), P(B))$ lies beneath a concave function similar to the function labeled $maxP(B)$ in the Figure. Its lower bound would always be 0. Finally in region **V**, $(P(A), P(B))$ pairs are bounded below a convex function similar to that shown as $minP(B)$.

Figure 2 displays examples of the bounds on $(P(A), P(B))$ in Region **II** in six cases where $\beta = \alpha > 1$. An indicator function value of 1 is displayed for the probability pairs cohering with several values of α from 1.1 through 4. It is evident that as $\beta = \alpha$ increases without bound, the non-linear restrictions on $(P(A), P(B))$ converge toward $P(A) + P(B) \leq 1$.

It is also interesting to view how distributions associated with the five regions of the parameter space fill the unit-simplex of all bivariate distributions over A and B. Figures 3, 4 and 5 display arrays of distributions representable by an (α, β) pair within each of the regions.

4. One implication of the bounding functions on $(P(A), P(B))$ pairs is that exchangeable assessments of A and B are representable by (α, β) pairs in any region *except* region **V** which precludes exchangeability because it does not allow $P(A) = P(B)$. (Remember that exchangeability of two events is equivalent to their equiprobability.) However, the *assertion* of exchangeability *in addition to* the restriction that $P(AB) = P(A)^\alpha P(B)^\beta$ would *require* that $P(A) = P(B)$. When this is the case, $P(AB)$ is constant over all (α, β) pairs whose sum is identical. Thus, exchangeable distributions can be represented and studied under the restriction that $\beta = \alpha$. The infinitely extendible exchangeable distributions are represented by the specification $\alpha = \beta \in [\frac{1}{2}, 1]$, whereas the finitely extendible exchangeable distributions are those for which $\alpha = \beta > 1$.

Geometrically, the exchangeable distributions lie on a plane within the unit-simplex, according to the restriction that $P(A) = P(B)$. Equivalently, $p_{01} = p_{10}$. (See the analysis of Diaconis (1977) as discussed in Lad (1996).) Figure 6 displays this plane, showing curves of distributions representable by a few selected parameter configurations. Parameter specifications in region **I**, where $\alpha + \beta \leq 2$, exhaust the infinitely extendible exchangeable distributions, while those in region **II**, for which $\alpha + \beta > 2$ exhaust the finitely extendible distributions.

We have been displaying aspects of the array of probability distributions that are representable by the assertion $P(AB) = P(A)^\alpha P(B)^\beta$ via restrictions on the cohering pairs $(P(A), P(B))$ it implies. The reverse analysis is also enlightening. We turn to a comment on the array of (α, β) pairs that can represent a specified coherent distribution over A and B.

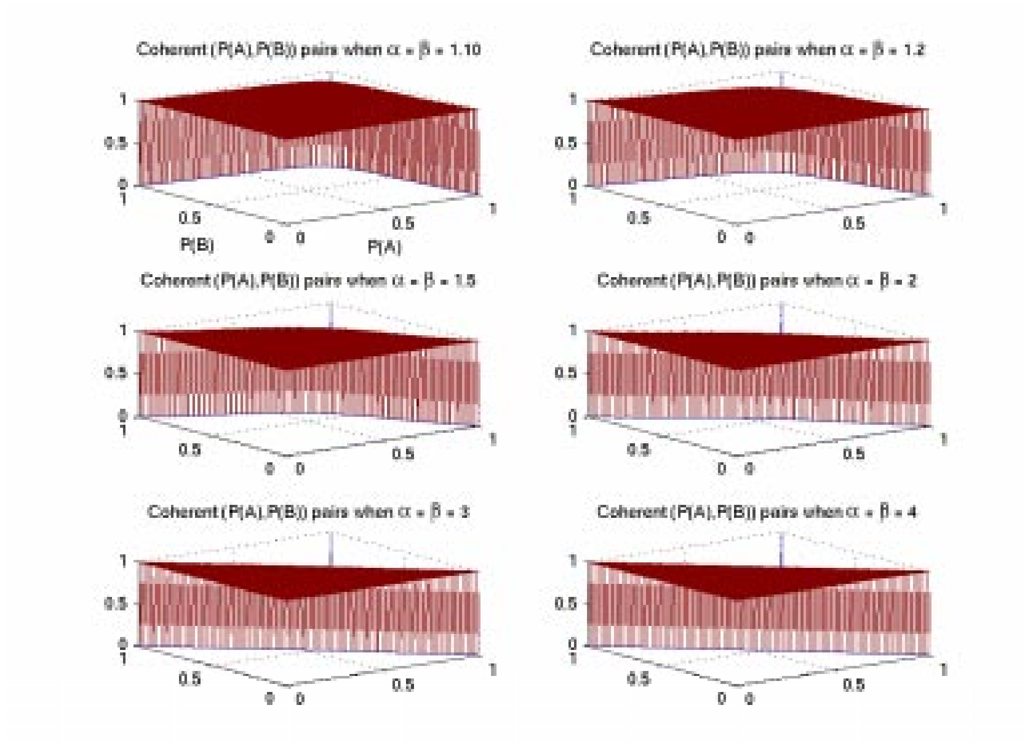

Figure 2: In each of the plots, values of $(P(A), P(B))$ pairs cohering with the specified values of $\alpha = \beta$ are indicated by function values of 1 over subregions of the unit-square. As $\alpha = \beta \to \inf$, the bound on $(P(A), P(B))$ converges toward $P(A) + P(B) \leq 1$. As shown, this bound is virtually reached by the time $\alpha = \beta = 2$.

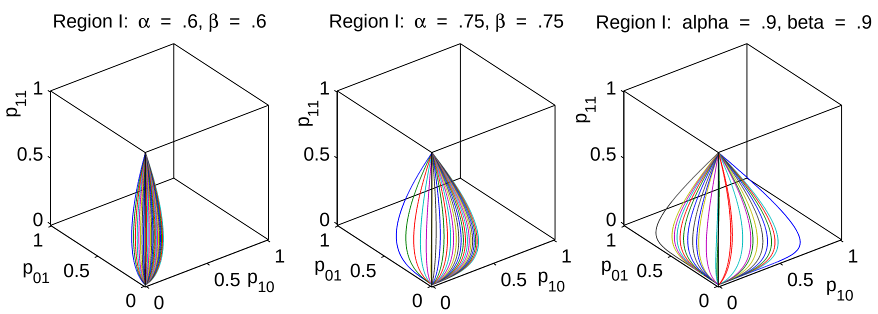

Figure 3: Displays of mass functions representable by three different configurations of $\alpha = \beta$ in region I.

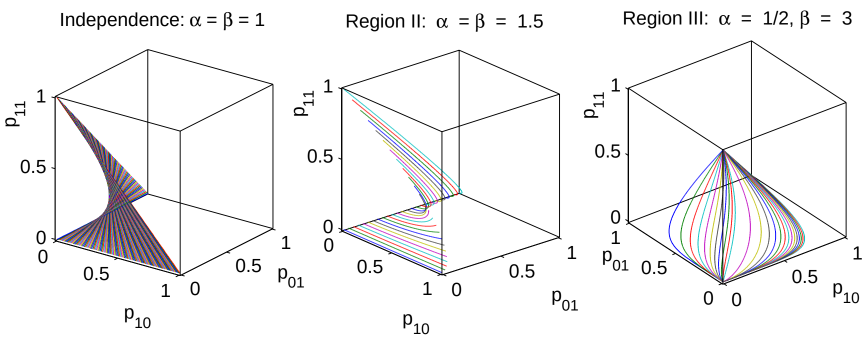

Figure 4: Displays of mass functions representable by independent distributions, by one configuration of $\alpha = \beta$ in region II, and one configuation in III.

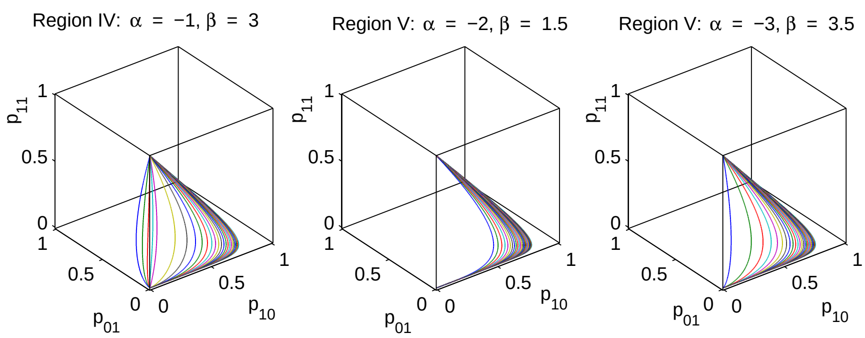

Figure 5: Displays of mass functions representable by one configuration of (α, β) in region IV, and by two configurations in region V.

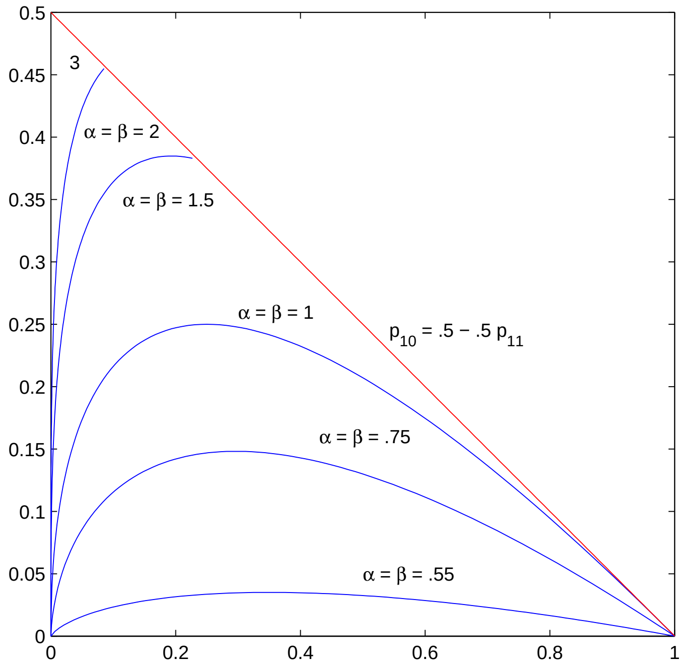

Figure 6: The plane of exchangeable probability distributions within the unit-simplex is filled by lines of distributions associated with (α, β) pairs within regions **I** and **II** along with the additional restriction that $\alpha = \beta$. Parameter specifications in region **I**, where $\alpha = \beta \in [1/2, 1]$, exhaust the infinitely extendible exchangeable distributions. Those in region **II**, where $\alpha = \beta > 1$, exhaust the finitely extendible exchangeable distributions.

5. Denote again by $(p_{00}, p_{01}, p_{10}, p_{11})$ a vector of coherent probability assertions for the events $P(\tilde{A}\tilde{B}), P(\tilde{A}B), P(A\tilde{B})$ and $P(AB)$, which identifies a probability distribution over the events A and B. Taking logarithms of the specification $P(AB) \;=\; P(A)^{\alpha}\ P(B)^{\beta}$ yields an equation for all (α, β) pairs that can represent this distribution, viz.,

$$\beta \quad = \quad \frac{ln\ (p_{11})}{ln\ (p_{01} + p_{11})} \quad - \quad \frac{ln\ (p_{10} + p_{11})}{ln\ (p_{01} + p_{11})}\ \alpha \quad \equiv \quad c \ - \ m\ \alpha \qquad (1)$$

The intercept coefficient of these equations, c, is at least equal to 1, whereas the slope coefficient, m, need only be positive. (Notice the negative sign on this positive coefficient in equation 1.) The slope coefficient, m, equals 1 only when $p_{01} \;=\; p_{10}$, identifying the exchangeable distributions. In these cases, the specific associated exchangeable distributions are identified by the value of the intercept, c. Figure 7 displays three lines of (α, β) pairs that support a specific distribution.

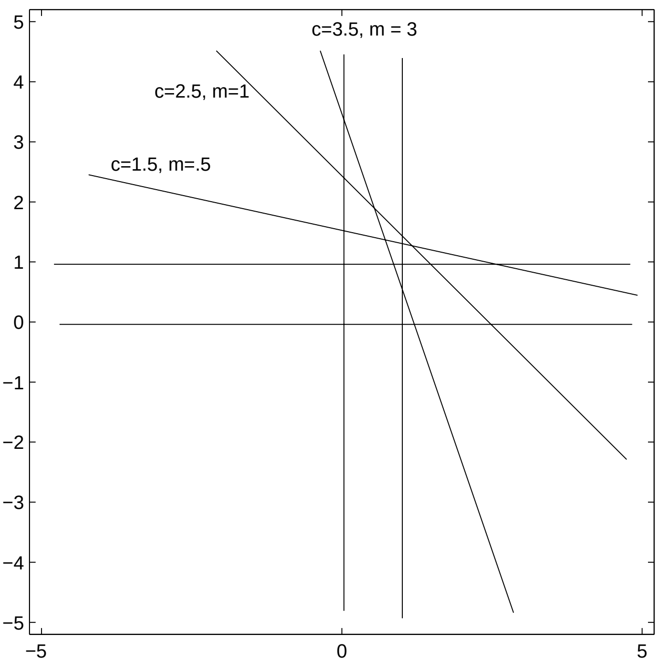

Figure 7: Three lines of (α, β) pairs that support specific distributions.

6. Our final concern are the conditional probabilities associated with the specification that $P(AB) \;=\; P(A)^{\alpha}\ P(B)^{\beta}$. The conditional probabilities have the forms

$$P(B|A) \;=\; P(A)^{\alpha-1}\ P(B)^{\beta} \qquad \text{and} \qquad P(A|B) \;=\; P(A)^{\alpha}\ P(B)^{\beta-1}$$

along with

$$P(B|\tilde{A}) \;=\; \frac{P(B)-P(A)^{\alpha}P(B)^{\beta}}{1-P(A)} \qquad \text{and} \qquad P(A|\tilde{B}) \;=\; \frac{P(A)-P(A)^{\alpha}P(B)^{\beta}}{1-P(B)}$$

for appropriate values of P(A) and P(B). (Remember that coherent joint assertions of P(A) and P(B) are restricted according to the values of α and β.) For examples, $\alpha \;=\; \beta \;=\; \frac{1}{2}$ implies $P(B|A) \;=\; 1$ and $P(A|B) \;=\; 1$ since it is required under this configuration that $P(A) = P(B)$. On the other hand, as is well-known, $\alpha = \beta = 1$ implies $P(B|A) = P(B)$ and $P(A|B) = P(A)$, the case of independence. Graphical depictions of conditional probabilities $P(B|A)$, $P(B|\tilde{A})$ and $P(A|B)$, $P(A|\tilde{B})$ for allowable P(B) values when $\alpha \;=\; \beta \;=\; \frac{3}{4}$ and $P(A) = .4$ appear in Figures 11 and 12.

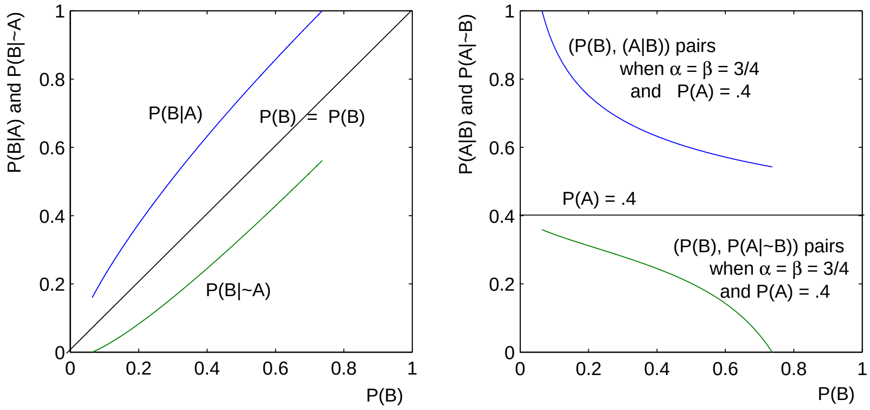

Figure 11. Relations of P(B|A) and P(B|~A) to P(B) over the range of cohering P(B) values when α = β = 3/4 and P(A) = .4. Cohering values of P(B) must lie within [.054, .736]. The line P(B) = P(B) appears only for reference.

Figure 12. P(A|B) and P(A|~B) values for the range of cohering P(B) values when α = β = 3/4 and P(A) = .4. Again the values of P(B) must lie within [.054, .736] as in Fig. 11. The line P(A) = .4 appears only for reference.

Figure 8: Notice the odd figure numbering here. These two figures are called Figures 11 and 12 in the text.

3 Concluding comments

Bounds on probabilities have been proposed as appropriate for statistical modelling in writings as diverse as Shafer (1976), Walley (1991) and Lad (1996) who exposits de Finetti's fundamental theorem of probability. One of the features of the factorisation representation we have have been studying is that it provides a convenient format for introducing bounds into stochastic modelling in specific forms.

REFERENCES

Diaconis, P. (1977) Finite forms of de Finetti's theorem of exchangeability, *Synthese*, **36**, 271-281.

Lad, F. (1996) *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley.

Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton: Princeton University Press.

Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.