Revision of Marginal Probability Assessments

Peter Jones, Sanjoy Mitter
EECS Dept
Massachusetts Institute of Technology
Cambridge, MA, U.S.A.
jonep@ll.mit.edu, mitter@mit.edu

Venkatesh Saligrama
ECE Dept
Boston University
Boston, MA, U.S.A.
srv@bu.edu

Abstract – *Sets of experts are frequently called upon to assess the occurrance probability of uncertain events. Often, the events being assessed are structurally interrelated, levying constraints on the assessements. However, expert panels frequently violate those constraints, resulting in internally inconsistent probability assessments. Based on a formulation of coherence due to de Finetti, we investigate methods for optimally modifying expert assessments to match structural constraints on the joint probability distribution over the set of events.*

Keywords: Probability theory, coherence, distributed probability assessment.

1 Introduction

Consider the following two questions:

1. What is the probability the U.S. will advance to the knock-out round of the 2010 World Cup?

2. What is the probability that England or Spain will make it to the finals?

Analysis of these questions might be undertaken by a number of experts including oddsmakers, advertising agents, or sports ministers. Other fields have similar interest in the accurate assessment of uncertain events. Medical diagnostics, investment banking, opinion polling, military intelligence, and a host of other industries depend on the fusion of expert probability assessments.

Ideally expert opinion would accurately reflect the outcome of uncertain events. Unfortunately, experts often make assessments that are not only biased and inaccurate [7], but which cannot be logically reconciled, particularly when the events under assessment are interrelated in complex and subtle ways. Consider again the sports example given above: the two probabilities may be treated independently, but in fact they are inherently related through the structure of the World Cup tournament and the rules of play and advancement. The requested probabilities are best viewed as marginals of a joint probability distribution. The advancement structure of the World Cup creates constraints on the set of valid joint distributions, constraints which aren't always intuitively obvious to experts when assessing uncertain outcomes.

There are numerous types of structural constraints that arise in the context of probability assessment. Events may be known to be Markov with respect to some graph, resulting in conditional independence constraints between random events. Events, such as the repeated flips of a coin with unknown bias, might be known to be exchangeable. Or they may have constraints due to set theoretic relationships between events. Constraints on the space of joint probabilities due to such relationships can potentially result in expert assessments of marginal probability that are inconsistent (or infeasible).

Given a set of irreconcilable probability assessments, the question arises of how best to revise the assessments to attain internal consistency.

1.1 Previous Work

Bruno de Finetti [2] formulated a consistency principle for probability assessments in the face of structural constraints due to set theoretic relationships. In de Finetti's formulation, a set of assessments is *coherent* if, treating the probability assessments as gambling odds, there is no wager that has guaranteed positive payout. It can be shown that this set of coherent assessments is convex.

In a series of papers [8, 6, 4] coherence is used to formulate an optimization-based method for aggregating experts' assessments of probability. In [8, 6] the suggestion is to select as the fused probability assessment of a set of experts the point in the coherent that lies closest

This work was sponsored by the U.S. Government under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government

(in terms of the standard Euclidean norm) to the vector of expert assessments. This is termed the *Coherent Approximation Principle (CAP)*. In [4] an approximation technique is suggested to deal with the potential combinatorial growth (in the number of assessed events) of the computation of the exact CAP solution.

1.2 Contributions of this Paper

Our contributions in this paper are threefold. First, we use the optimization framework of [8] to find a class of characteristic matrices for which the CAP solution has a convenient closed form. We then extend this solution to a larger class of characteristic matrices for which the solution, while not exact, is boundedly suboptimal and provably coherent.

Second, we suggest possible limitations in the coherent approximation formulation due to the use of the Euclidean norm as the objective function and introduce a information divergence-based objective function that overcomes some of the limitations.

Third, we investigate the effects of combining coherence constraints with constraints due to exchangeability and Markovianity. We find conditions under which these additional constraints do not impact the space of coherent marginal probability assessments, independent of the specific objective function employed.

2 Coherent Approximation

De Finetti's definition of coherence as a set of odds that admit no guaranteed positive payoff (see Section 1.1) also has a useful geometric interpretation. To explain this interpretation, we introduce some mathematical notation that will be used throughout the paper. Let $\Omega = \{\omega_1, \omega_2, \ldots \omega_M\}$ be a finite event space, and let $\mathcal{A} = \{A_1, A_2, \ldots, A_N\}$; $A_i \subseteq \Omega$. Note that $\mathcal{A}$ is not necessarily an algebra on the space Ω. Let $P \in [0,1]^N$ be a probability assessment over the events $\{A_i\}$ and let $\chi \in \{0,1\}^{N \times M}$ defined by

$$\chi_{ij} = \begin{cases} 1 & \omega_j \in A_i \\ 0 & \text{otherwise} \end{cases}$$

be the characteristic matrix for the set of events. Then, following section 3.4 of [2], a probability assessment is coherent if and only if $\exists \lambda \in [0,1]^M$ with $\sum_i \lambda_i = 1$ s.t. $P = \chi\lambda$. If such a λ exists, we will say $P \in \texttt{convhull}(\chi)$.

2.1 Coherence Example

Consider the following assessment space:

$$\begin{aligned} \Omega &= \{\omega_0, \omega_1, \ldots, \omega_6\} \\ A_1 &= \{\omega_1, \omega_3, \omega_5\} \\ A_2 &= \{\omega_2, \omega_3, \omega_6\} \\ A_3 &= \{\omega_4, \omega_5, \omega_6\} \end{aligned}$$

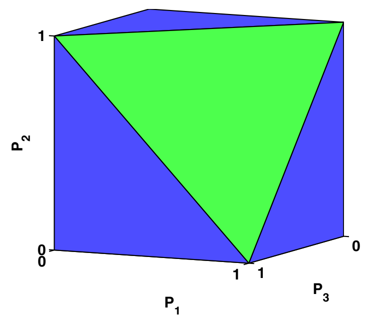


Figure 1: The set of coherent marginal probabilities generated by χ

The characteristic matrix can been seen to be:

$$\chi = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

and the set of coherent assessments, $\mathcal{P}$, is shown in Figure 1.

2.2 Coherent Approximation Principle

Osherson and Vardi [8] use the concept of coherence to formulate an optimization problem for the fusion of a panel of assessors. Given a panel's assessment P the optimal fused assessment $P^* = \chi\lambda^*$ is defined by

$$\lambda^* = \text{argmin}_{\{\lambda | \sum_i \lambda_i = 1,\ \lambda_i \geq 0\}} ||P - \chi\lambda||_2 \qquad (1)$$

Note that the optimization occurs in the space of atomic events Ω, which may grow exponentially in the assessment space size. To mitigate this computational challenge, previous authors [4] proposed a hybrid approach between linear averaging and coherent approximation. Unfortunately, this approach will generally produce a fused estimate that is not coherent.

In the following subsections we develop a general fusion rule that operates in the assessment space and generates a coherent fused assessment.

2.3 Monadic Structure

Consider the class of characteristic matrices such that

$$\sum_j \chi_{ij} \leq 1$$

We will refer to matrices in this class as *monadic*, meaning that each event under assessment is, at most, a singleton.

In this case, it is simple to show that a closed-form solution exists to the problem of finding a coherent approximation to an incoherent assessment. Let $P^* = \chi\lambda^*$ be defined by (1) and let

$$\bar{P}_i = \frac{1}{n_i} \sum_{j \in \mathcal{N}_i} P_j$$

where $\mathcal{N}_i = \{j | A_j == A_i\}$ and $n_i = |\mathcal{N}_i|$. Let the probability excess/deficit be denoted as $D = 1 - \sum_{j=1}^{N} \frac{1}{n_j} \bar{P}_j$ and assume wlog that $n_1 \bar{P}_1 \leq n_2 \bar{P}_2 \leq \ldots \leq n_N \bar{P}_N$ and $\forall i \ \mathcal{N}_i, \ j < k, \ j, k \in \mathcal{N}_i \Rightarrow \{j, j+1, \ldots, k\} \subseteq \mathcal{N}_i$.

Theorem 1 *If χ is monadic, then*

$$P^* = \bar{P} + \Delta$$

where we define $\Delta \in [0,1]^N$ as

$$\Delta_i = \begin{cases} f(\{\Delta_j\}_1^{i-1}, \{n_j\}, \bar{P}) & D < 0 \\ \frac{D}{n_i} \left(\sum_j \frac{1}{n_j^2} \right)^{-1} & D \geq 0 \ \& \ \nexists n_k = 0 \\ 0 & \text{o.w.} \end{cases}$$

with, for a given $i \in \{1, 2, \ldots, N\}$,

$$f(\cdot, \cdot, \cdot) = \begin{cases} \Delta_{i-1} & \mathcal{N}_i = \mathcal{N}_{i-1} \\ \max \left(-\bar{P}_i, \frac{n_{i-1}}{n_i} \Delta_{i-1} + g_i \right) & \text{o.w.} \end{cases}$$

with $g_i = \frac{1}{n_i} \left(\sum_{j=i}^{N} \frac{1}{n_j^2} \right)^{-1} \left(D - \sum_{j=1}^{i-1} \frac{\Delta_j}{n_j} \right)$ and $\Delta_0 = \bar{P}_0 = 0$. The proof of this theorem is omitted here in the interest of space.

Theorem 1 can be extended to a broader class of matrices by using the complementary completeness of probabilities (i.e. each assessment P of A is also an indirect assessment $(1 - P)$ of $A^c = \Omega \backslash A$). A characteristic matrix is said to be *generalized* monadic if

$$\sum_j \chi_{ij} \leq 1 \qquad \text{or} \qquad \sum_j \chi_{ij} \geq N - 1$$

As before, let P^* be defined by (1). Let $\bar{P}_i$ be the mean of all assessments of event A_i (both directly over A_i and indirectly over A_i^c).

Corollary 1 *If χ is generalized monadic, then*

$$\forall i \text{ s.t.} \sum_j \chi_{ij} \leq 1, P_i^* = \bar{P}_i + \Delta_i$$

Also

$$\forall i \text{ s.t.} \sum_j \chi_{ij} \geq N - 1, P_i^* = 1 - \bar{P}_i + \Delta_i$$

with Δ defined analogously to Theorem 1.

The benefit of Theorem 1 and its generalization is that there exist certain matrices for which the CAP can be solved exactly, with computational complexity proportional to the number of events under assessment (rather than the potentially exponentially larger number of atomic events, as in the direct solution to the CAP problem).

2.4 Suboptimal approximation

There is not a simple extension of the result from Section 2.3 to generally structured characteristic matrices. However, it is possible to use the solution for monadic matrices to approximate the solution for general characteristic matrices.

Consider again the coherence constraint: $P = \chi\lambda$. This can be rewritten in the following way

$$P = \sum_j [\chi_{i,S_j}] \lambda_{S_j}$$

where $\{S_j\}$ forms a partition over $\{1, 2, \ldots, |\Omega|\}$. Essentially, we've decomposed the original characteristic matrix columnwise. It is simple to show that for any characteristic matrix there exists a columnwise decomposition such that $\forall j \ [\chi_{i,S_j}]$ is generalized monadic.

Also, Theorem 1 can be generalized to the constraint $\sum \lambda_i = \alpha_j$ where α_j is some given constant, rather than $\sum \lambda_i = 1$. Combining these two results gives a method of suboptimal coherent approximation

1. Decompose characteristic matrix columnwise into a set of monadic matrices
2. Apply Theorem 1 to each subproblem, with the constraint that $\sum_i (\lambda_{S_j})_i = \alpha_j$
3. Using Lagrangian analysis, determine optimal α_j coupling constants

The resulting solution is guaranteed coherent. Furthermore, the suboptimality of the result can be bounded by $\sum_i \sum_{j \neq i} \lambda_{S_i}^T \lambda_{S_j}$.

3 Divergence-based Costs

In the formulation of CAP, a Euclidean objective function was assumed without justification. In [5], a more general formulation of the CAP problem is considered and it is demonstrated that the Euclidean objective function in (1) is not inherent. In the sequel we will make two arguments for an objective function based on binary information divergence rather than quadratic cost. In Section 3.1 we provide a mathematical model of expert assessment that imlies a binary information divergence-based objective function. Then, in Section 3.2 we illustrate the superiority of this objective function in revising expert assessments when χ is the identity (and hence P must be a probability mass function to be coherent).

3.1 Opinion deformation and Sanov's Theorem

Consider the following model for probability assessors: each assessor observes a sequence of realizations of her event over several periods of time and maintains an empirical distribution of event occurrance. When called upon to make an assessment, the assessor selects a distribution approximately equal to her empirical distribution from finite set $\mathcal{P}$.

When all assessors have reported it is noted the set of assessments is incoherent and therefore at least one assessor is in error, either due to approximation or due to miscalculation of the empirical distribution. Given that at least one reported assessment is in error, which is the most likely generating distribution to have caused the error(s)?

First, let's consider the most likely estimate of a true distribution when we know a particular assessor is in error. Let p^* be the true generating distribution for the erroneous assessor, and $\hat{p}$ be the reported distribution. From Sanov's theorem, the probability of erroneously declaring p_j when the true generating distribution is p_i decays exponentially in n and is approximately

$$p(\hat{p} = p_j | p^* = p_i, p^* \neq \hat{p}) \doteq \exp(-nD_b(p_j || p_i))$$

where

$$D_b(p_j||p_i) = p_j \log\left(\frac{p_j}{p_i}\right) + (1-p_j)\log\left(\frac{1-p_j}{1-p_i}\right) \quad (2)$$

and $\doteq$ denotes asymptotic equality to within a multiplicative factor with a slower rate of decay. (More accurately, the asymptotic rate of decay is $\inf_{p\in A(p_j)} D_b(p \,\|\, p_i)$ where $A(p_j)$ is the set of all distributions mapped to p_j by $\hat{p}$. We've made the simplifying assumption that the acceptance regions $A(p_j)$ are uniformly small in a divergence sense).

Now, assume the prior probability that $p^* = p_i$ is uniform over all p_i; then the joint distribution $p(\hat{p} = p_j, p^* = p_i | p^* \neq \hat{p})$ is (asymptotically, approximately) proportional to $\exp(-nD_b(p_j||p_i))$.

Next, given that $\hat{p} = p_j$, the conditional distribution $p(p^* = p_i | \hat{p} = p_j, p^* \neq \hat{p})$ is also proportional to $\exp(-nD_b(p_j||p_i))$. Therefore the maximum likelihood estimate of p^* given $\hat{p} = p_j$ and $\hat{p} \neq p^*$ is asymptotically

$$p_i^* = \operatorname*{argmin}_{p_i \in \mathcal{P}\setminus\{p_j\}} D_b(p_j||p_i) \quad (3)$$

Moving to the multiple assessor case, given a set of incoherent assessments $p_1, p_2, \ldots, p_N$ we know at least one of the assessments is in error. If we knew which one, by the preceding development it should be revised following Equation 3. Since we don't know which assessment (or combination of assessments) was in error, heuristically we can distribute risk equally across all distributions, leading to a reformulation of the CAP as

$$\lambda^* = \operatorname{argmin}_{\{\lambda | \sum_i \lambda_i = 1,\ \lambda_i \geq 0\}} \sum_{i=1}^{N} D_b(p_i \,\|\, \chi_i \lambda) \quad (4)$$

where χ_i is the i^{th} row of χ.

3.2 CAP for Probability Mass Funtions

Consider the case when $\chi = I$ (the identity). In this case, we can rewrite the optimization problem under a general cost function $C(P, Q)$ as

$$Q^* = \operatorname{argmin}_{\{Q | \sum_i Q_i = 1,\ Q_i \geq 0\}} C(P, Q) \quad (5)$$

3.2.1 Solution under Quadratic Cost Function

Since the identity is a monadic matrix (see Section 2.3), if we take $C(P, Q)$ to be the quadratic cost function then the CAP solution is given by Theorem 1. The solution attempts to deform each element in the assessment vector by an equal amount, while respecting the range constraints on the probabilities.

Example: Solution under Quadratic Cost As an example, take

$$P = \begin{bmatrix} 0.2 \\ 0.7 \\ 0.7 \end{bmatrix}$$

By Theorem 1 the solution to (5) is

$$Q^* = \begin{bmatrix} 0.0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

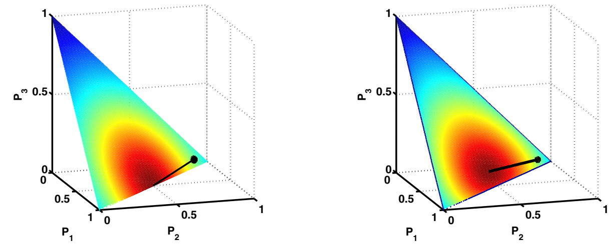
This example demonstrates one of the limitations of using quadratic cost. The change in terms of subjective belief from 0.2 to 0.0 is greater than the change from 0.7 to 0.5. Cognitively, the leap from moderate uncertainty (0.2) to absolute certainty (0.0) is larger than between moderate uncertainty (0.7) and somewhat higher uncertainty (0.5). The cognitive difference between believing an event is 20% likely and believing the same event is 0% likely is very large!

A more natural method for transforming a set of assessments into a probability mass function is simply to scale the assessments until they sum to one. In the following section it will be seen how the binary information divergence cost explains both why this is natural and why it is not the right thing to do (quite).

3.2.2 Solution under Divergence-based Cost

Take $C(P, Q) = \sum_{i=1}^{N} D_b(P_i||Q_i)$ where $D_b(P_i||Q_i)$ is given by Equation 2. The rationale for choosing this cost function is given in Section 3.1. The projection under both the Euclidean and the binary information divergence cost functions for the example given in Section 3.2.1 is shown in Figure 2. Of particular note is the natural barrier created by the divergence-based cost function at the boundaries of the simplex. This reflects the intuitive result that revision to absolute certainty along any dimension is very costly.

Unfortunately, even with $\chi = I$, the analysis for this cost function is difficult. We can, however, derive a simple lower bound on the cost. Let $C_1(P, Q) = \sum_{i=1}^{N} P_i \log\left(\frac{P_i}{Q_i}\right)$ and $C_2(P, Q) =$



(a) (b)

Figure 2: Optimal coherent revisions for example under (a) quadratic cost and (b) binary information divergence cost

$\sum_{i=1}^{N}(1-P_i)\log\left(\frac{1-P_i}{1-Q_i}\right)$. Then $C(P,Q)=C_1(P,Q)+C_2(P,Q)$ and

$$C(P,Q^*) \geq C_1(P,Q_1^*) + C_2(P,Q_2^*)$$

where $Q_i^* = \text{argmin}_{\{Q|\sum_i Q_i=1,\ Q_i\geq 0\}} C_i(P,Q)$ and Q^* is defined by (5).

Using Lagrangian analysis, we can solve explicitly for Q_1^*. Specifically,

$$Q_1^* = \frac{P}{\sum_i P_i}$$

The optimal projection under cost $C_1(P,Q)$ is a scaling of the assessments to form a probability mass function. This is the "natural" solution suggested at the end of Section 3.2.1.

To derive the other half of the lower bound, let

$$\hat{Q}_2^* = \underset{\{Q|\sum_i Q_i=1\}}{\text{argmin}} \quad C_2(P,Q)$$

be the solution to the projection while neglecting the range constraints on Q. Obviously $C_2(P,Q_2^*) \geq C_2(P,\hat{Q}_2^*)$. By analysis symmetric to the case for Q_1^*, we find

$$1-\hat{Q}_2^* = \frac{1-P}{\sum_i 1-P_i}$$

The optimal projection, neglecting range constraints, under cost $C_2(P,Q)$ is a scaling of the implicit assessments against the complementary events $\{A_i^c\}$.

Lemma 1 *$C(P,Q^*)$ is bounded from below by*

$$C(P,Q^*) \geq C_1\left(P,\frac{P}{\sum_i P_i}\right) + C_2\left(P,1-\frac{1-P}{\sum_i 1-P_i}\right)$$

The derivation of the bound above gives insight into why the natural inclination to simply normalize the assessments isn't quite the right thing to do. Each expert's assessment P_i specifies not only the subjective probability of event A_i, but also the subjective probability of event A_i^c. Taking $Q=\frac{P}{\sum_i P_i}$ neglects the implicit assessed probabilities of the complementary events. In using $C(P,Q)=\sum_{i=1}^{N} D_b(P_i||Q_i)$ as the cost function for (5), the cost of normalizing the assessments is balanced against the cost of normalizing their complements.

4 Exchangeability

Regardless of what utility function is employed in CAP, the constraint set is determined by the convex hull of the columns of the characteristic matrix. The question we address in the next two sections is how adding structural constraints to the problem effects the solution, independent of the specific choice of objective function.

In addition to the coherence principle de Finetti also introduced the concept of *exchangeability* of random variables. In one sense, exchangeability is a generalization of the concept of independent and identically distributed. However, more relevant to our development is the constraint it induces on the joint distribution of the random variables.

A set of random variables is said to be *exchangeable* if it is invariant under permutation. Let τ be a one-to-one mapping $\tau:\{1,2,\ldots,N\}\rightarrow\{1,2,\ldots,N\}$. Then, by definition, a set of random variables is exchangeable if $P(X_{\{1,2,\ldots,N\}}) = P(X_{\tau(\{1,2,\ldots,N\})})$ for all τ.

If the only structural information about a set of binary random variables is that they are exchangeable, then the only constraint on the joint probability distribution is that it is symmetric (in the sense of invariance to permutation). As a consequence of this symmetry, it is simple to show that the marginal probabilities of a set of exchangeable random variables must be equal. Indeed, this is the only constraint levied on the marginal probabilities by exchangeability. In other words, given a set of binary random variables $X_1, X_2, \ldots, X_n$, exchangeability implies $P(X_i)=P(X_j)$, $\forall i,j$. There are no other implications with regards to the marginal probabilities. So, consistent with the geometric interpretation of coherence in Section 2.1, one can consider the set of coherent marginals with respect to exchageability to be those that lie on the line connecting the $[0]^n$ and $[1]^n$ vertices of the hypercube.

4.1 Matched exchangeability constraints

When multiple forms of structural information are known about a set of random variables, a consistent estimate of the probabilities must be in agreement with all the constraints. When a set of random variables are known to be 1) exchangeable and 2) coherent w.r.t. a characteristic matrix, the set of consistent probabilities is restricted more than by taking either structural requirement individually.

There are situations in which the characteristic matrix can be "matched" to the exchangeability constraint, i.e. where coherence levies no constraints above and beyond exchangeability.

Consider a characteristic matrix χ . Assume w.l.o.g. that all columns are unique and that $|\Omega| \leq 2^N$. Let $n_i = \sum_j \chi_{ij}$, and consider sets $J_k = \{i | n_i = k\}$. Note that $|J_k| \leq \begin{pmatrix} N \\ k \end{pmatrix}$. The following lemma is a direct consequence of the invariance under permutation property of exchangeable random variables.

Lemma 2 *If $\mathcal{A}$ is a set of exchangeable binary random variables with characteristic matrix χ and marginal probabilities $P = \chi\lambda$ then $\forall i, j,\ n_i = n_j \Rightarrow \lambda_i = \lambda_j$ and $\forall k,\ |J_k| < \begin{pmatrix} N \\ k \end{pmatrix} \Rightarrow \lambda_i = 0\ \forall i$ s.t. $n_i = k$.*

An immediate corollary of Lemma 2 gives necessary and sufficient conditions under which a characteristic matrix is matched with exchangeability.

Corollary 2 *A characteristic matrix χ is matched to the exchangeability constraint iff $|J_k| = 0$ for all $k \notin \{0, N\}$*

Another corollary gives conditions on a characteristic matrix such that the set of feasible marginals under both exchangeability and a characteristic matrix is exactly the intersection of the feasible sets under each structural condition taken independently.

Corollary 3 *For event set $\mathcal{A}$ let $\mathcal{P}_1$ be the set of marginals consistent with exchangeability, $\mathcal{P}_2$ be the set of marginals consistent with a characteristic matrix and $\mathcal{P}$ be the set of marginals consistent with both exchangeability and a characteristic matrix. Then $\mathcal{P} \subseteq \mathcal{P}_1 \cap \mathcal{P}_2$ with equality iff $|J_k| \in \left\{0, \begin{pmatrix} N \\ k \end{pmatrix}\right\}$ for all $k \in \{0, 1, ..., N\}$*

An example of an event set that doesn't meet the necessary requirements for equality from Corollary 3 is given in Section 4.2.

4.2 Non-additivity of constraints

As stated earlier, if the members of a set of random variables are exchangeable, the feasible set of marginal probabilities are exactly those which lie on the line segment between the origin and unity. To demonstrate the non-additivity of the marginal constraints, consider the following example (closely related to the example in Section 2.1):

$$\begin{aligned} \Omega &= \{\omega_0, \omega_1, \ldots, \omega_6\} \\ A_1 &= \{\omega_1, \omega_3, \omega_5\} \\ A_2 &= \{\omega_2, \omega_3\} \\ A_3 &= \{\omega_4, \omega_5\} \end{aligned}$$

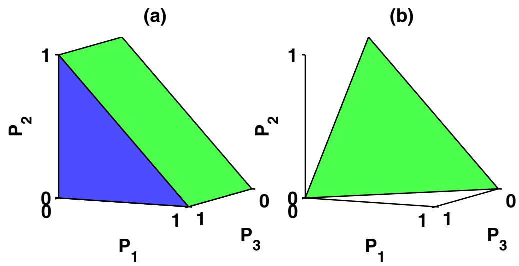


Figure 3: (a) Feasible marginals based only on characteristic matrix (b) Feasible marginals after including exchangeability constraints (omitting the equality constraint for expositional clarity)

Its characteristic matrix is

$$\chi = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

and the set of coherent marginals is shown in Figure 3(a). Note from the assessment space that

$$A_1^c \cap A_2 \cap A_3 = \emptyset \Rightarrow P(A_1^c \cap A_2 \cap A_3) = 0$$

Now, if we assume additionally that the random variables are exchangeable, then

$$P(A_2^c \cap A_3 \cap A_1) = P(A_3^c \cap A_1 \cap A_2) = 0$$

This is equivalent to the requirement that $\lambda_3 = \lambda_5 = 0$. In addition, as stated earlier, exchageability requires that the marginals all be equal. These two constraints together imply that only marginals of the form $P = [\alpha, \alpha, \alpha]$ where $0 \leq \alpha \leq \frac{1}{3}$ are consistent with the structural constraints. Note, however, that the intersection of the feasibility sets for the two types of structural constraints taken independently is $P = [\alpha, \alpha, \alpha],\ 0 \leq \alpha \leq \frac{\sqrt{2}}{2}$. Because the conditions of Corollary 3 are not satisfied, the set of feasible marginals for a combination of two types of structural constraints is a strict subset of the intersection of the sets of feasible marginals of the structural constraints taken individually.

5 Markovianity

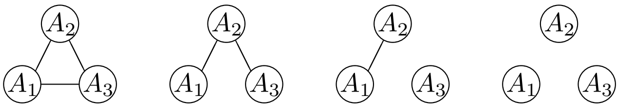
Just as in Section 4 we examined the combination of a characteristic matrix with exchangeability among the random variables, in this section we will investigate the combination of a characteristic matrix with Markov relationships among the random variables. To motivate the problem, consider the following situation

A group of individuals under suspicion of terrorist activities. Based on a variety of information sources, a group of experts estimate the probability that subsets of suspects are, in fact, involved in terrorist activity. Now, assume there is a known social network connecting the suspects. Social ties are a central element in fomenting terrorist activity, so much so that it can be assumed that the probability an individual is involved in terrorist activity is conditionally independent of the activities of all other individuals in the network, given his group of immediate social connections (his neighborhood). This is exactly equivalent to saying the participation of an individual in terrorist activities is Markov with respect to the graph representing the social network.

Let $G = (V, E)$ be a graph such that for every i, node v_i corresponds to A_i. Constrain the joint probability over $\mathcal{A}$ to be Markov w.r.t. G and let the set of realizable marginal probabilities be denoted $\hat{\mathcal{P}}$. The following example illustrates how constraining the joint probability to be Markov w.r.t. G can affect the set of feasible marginal probabilities.

5.1 Non-Additivity of Constraints revisited

Consider the assessment space from Section 2.1 with the set of coherent marginals shown in Figure 1 (and replicated in Figure 5(a)). When the marginals are Markov w.r.t. a complete graph (e.g. Figure 4(a)), the set of feasible marginals is exactly the coherent set.



(a) (b) (c) (d)

Figure 4: Four Markov graphs

Now, consider the graph shown in Figure 4(b). The additional constraint implied by this graph is $P(A_1 \cap A_3|A_2) = P(A_1|A_2)P(A_3|A_2)$. Since there does not exist $\omega \in \Omega$ s.t. A_1, A_2, and A_3 all obtain, $P(A_1 \cap A_3|A_2) = 0$. This implies that

$$P(A_1 \cap A_3|A_2) = 0 = \frac{\lambda_3}{\lambda_2 + \lambda_3 + \lambda_6}\frac{\lambda_6}{\lambda_2 + \lambda_3 + \lambda_6}$$

The additional constraints introduced by the graph imply that either $\lambda_3 = 0$ or $\lambda_6 = 0$ (or both).

Let $\hat{\mathcal{P}}$ be the set of marginal probabilities that are 1) coherent w.r.t. characteristic matrix χ and 2) Markov w.r.t. graph G. Figure 5(a)-(d) show $\hat{\mathcal{P}}$ for the G given by Figure 4(a)-(d) and χ given previously.

5.2 Complete Event Sets

Definition 1 *The set of events $\mathcal{A}$, and its corresponding characteristic matrix χ, is said to be **complete** if for every $a \in \{0, 1\}^N$ $\exists j$ s.t. $a = \chi(\omega_j)$.*

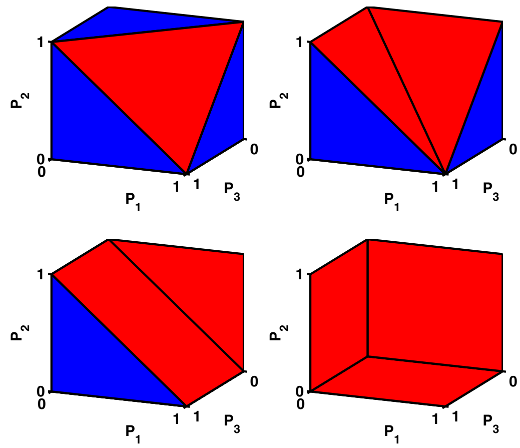


Figure 5: (a)Top-Left: $\mathcal{P}$, no graph constraints; (b) Top-Right: $\hat{\mathcal{P}}$ for graph in Figure 4(b); (c) Bottom-Left: $\hat{\mathcal{P}}$ for graph in Figure 4(c); (d) Bottom-Right: $\hat{\mathcal{P}}$ for graph in Figure 4(d)

For any complete $\mathcal{A}$, $\mathcal{P}$ is exactly the $N-$dimensional hypercube.

Theorem 2 *For any graph G, if $\mathcal{A}$ is complete then $\hat{\mathcal{P}} = \mathcal{P}$*

There exists a simple constructive proof for this theorem, omitted here in the interest of space.

Definition 2 *The set of events $\mathcal{A}$ and its corresponding characteristic matrix χ, is said to be **degenerately complete** if $\mathcal{A}$ can be decomposed into two disjoint sets $(\mathcal{A}_c, \mathcal{A}_d)$ where $\mathcal{A}_c$ is complete and $\mathcal{A}_d$ is deterministic (i.e. if $A_d \in \mathcal{A}_d$ then χ_{dj} is uniformly 0 or 1 for all j)*

Theorem 3 *For a general graph $G = (V, E)$, $\hat{\mathcal{P}} = \mathcal{P}$ iff, for each possible decomposition of the graph into $(\mathbb{P}, \mathbb{C}, \mathbb{F})$, where $\mathbb{C}$ is a cut set on G, the characteristic submatrices corresponding to each realization C of $\mathbb{C}$ are degenerately complete*

With the insight gained from Theorem 3, we can turn our attention to the central question of what effect constraining the joint probability space by graph G has on the coherence set $\mathcal{P}$. From de Finetti's theorem, we have $\mathcal{P} = \texttt{convhull}(\chi)$.

Suppose $(\mathbb{P}, \mathbb{C}, \mathbb{F})$ is a partition of graph G s.t. $\mathbb{C}$ is a cut set (and is not a superset of some other cut set). The set of instantiations of $\mathbb{C}$ form a partition on Ω. Let the subset of atomic events in an instantiation C of $\mathbb{C}$ be denoted Ω_C.

Now, suppose that for some instantiation C of cut set $\mathbb{C}$ of graph G, the conditions of Theorem 3 do not hold. Then, for the joint distribution to support the constraint implied by the graph, $\lambda_I = 0$ for some set

$I \subset \{0,1,\ldots,N\}$. Specifically, those atomic events which cause the degenerate completeness condition to be violated must receive zero weight.

Let $\mathcal{I}_C = \{I : \chi_{\mathcal{A}\backslash\mathbb{C}}(\Omega_C \backslash \omega_I)$ is degenerately complete$\}$ and let the elements of $\mathcal{I}_C$ be denoted I_C. Let $\mathcal{I}_{\mathbb{C}} = \left\{I : I = \bigcup_{C\in\mathbb{C}} I_C\right\}$ and let the elements of $\mathcal{I}_{\mathbb{C}}$ be denoted $I_{\mathbb{C}}$. Let $\mathcal{I} = \{I : I = \bigcup_{\mathbb{C}} I_{\mathbb{C}}\}$ and let $\hat{\mathcal{P}}_I = I(\mathcal{A},\Omega)\lambda$ s.t. $\lambda_I = 0$. Then

$$\hat{\mathcal{P}} = \bigcup_{I\in\mathcal{I}} \hat{\mathcal{P}}_I$$

Returning to the example, consider the graph shown in Figure 4(a). This can be decomposed into $\mathbb{P} = A_1$, $\mathbb{C} = A_2$, $\mathbb{F} = A_3$. First, conditioning on the event $\bar{A}_2 = \{\omega_0, \omega_1, \omega_4, \omega_5\}$, we get the characteristic submatrix

$$\chi_{\mathcal{A}\backslash\{A_2\}}(\bar{A}_2) = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

This submatrix is complete (and hence degenerately complete), and so there are no conditions on λ due to conditioning on the event $\bar{A}_2$ (i.e. $\mathcal{I}_{A_2=0} = \emptyset$).

Next, conditioning that event $A_2 = \{\omega_2, \omega_3, \omega_6\}$ results in the characteristic submatrix

$$\chi_{\mathcal{A}\backslash\{A_2\}}(A_2) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The characteristic submatrix is not degenerately complete. Therefore $\hat{\mathcal{P}} \neq \mathcal{P}$ (compare Figure 5(a) with Figure 5(b)). Since

$$\chi_{\{A_1,A_3\}}(A_2\backslash\omega_3) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\chi_{\{A_1,A_3\}}(A_2\backslash\omega_6) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

are degenerately complete, $\mathcal{I}_{A_2=1} = \{3,6\}$. The set I_{A_2} is therefore equal to $\{I : I = \bigcup_{i\in\{0,1\}} I_{A_2=i} = \{\{3,\emptyset\},\{6,\emptyset\}\} = \{3,6\}$. Since A_2 is the only cut set of G that is not a superset of another cut set, we have $\mathcal{I} = I_{A_2} = \{3,6\}$. Therefore the set $\hat{\mathcal{P}}$ can be described by

$$\hat{\mathcal{P}} = \bigcup_{i\in\{3,6\}} \{\chi\lambda | \sum_j \lambda_j = 1;\ \lambda_j \geq 0; \lambda_i = 0\}$$

6 Conclusion

Expert assessment of uncertain events impacts the quality of medical diagnoses, the strength of financial markets, the security posture of nations and innumerable other critical elements of modern life. Improving these assessments can have significant impact. By understanding the relationships between uncertain events and optimally combining the independent expert assessments, the overall performance of the assessment process can be improved.

Previous authors suggested an approximation principle to fuse assessments of events under coherence constraints. In this paper we have demonstrated a solution mechanism for the CAP problem, suggested a new objective function justified both theoretically and practically, and expanded the coherence formulation to a broader set of structural constraints.

There are many potential future avenues of research related to these results. Other interesting structural relations might be explored and the definition of coherence extended to them. Generalizing the concept of coherence from the set of marginal probabilities to a set of marginal *utilities* could relate these results to arbitrage and incoherent pricing or risk in markets. Some effort has already been made in this direction [3, 1]. Finally, extending this analysis to a multiple-period assessments would increase its applicability to many situations of practical importance.

References

[1] P. Artzner, F. Delbean, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, July 1999.

[2] Bruno de Finetti. *Theory of Probability*, volume 1-2. Wiley New York, 1974.

[3] F. Delbean. Coherent measures of risk on general probability spaces. In *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, pages 1–38. Springer–Verlag, 2002.

[4] J.B. Predd *et al.* Aggregating forecasts of chance from incoherent and abstaining experts. *Decision Analysis*, 5:177–189, 2008.

[5] J.B. Predd *et al.* Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792, October 2009.

[6] R. Batsell *et al.* Eliminating incoherence from subjective estimates of chance. In *8th Int'l Conf on the Principles of Knowledge Representation and Reasoning*, pages 353–364, Toulouse, France, 2002.

[7] Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.

[8] D.N. Osherson and M.Y. Vardi. Aggregating disparate estimates of chance. *Games and Economic Behavior*, 56(1):148–173, July 2006.