



Open Research Online

The Open University's repository of research publications and other research outputs

Conditionally externally Bayesian pooling operators in chain graphs

Journal Item

How to cite:

Faria, Álvaro and Smith, Jim Q. (1997). Conditionally externally Bayesian pooling operators in chain graphs. *The Annals of Statistics*, 25(4) pp. 1740–1761.

For guidance on citations see [FAQs](#).

© 1997 Institute of Mathematical Statistics

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

CONDITIONALLY EXTERNALLY BAYESIAN POOLING OPERATORS IN CHAIN GRAPHS

BY ÁLVARO E. FARIA, JR.¹ AND JIM Q. SMITH

University of Warwick

We address the multivariate version of French's *group decision problem* where the m members of a group, who are jointly responsible for the decisions they should make, wish to combine their beliefs about the possible values of n random variables into the group consensus probability distribution. We shall assume the group has agreed on the structure of associations of variables in a problem, as might be represented by a commonly agreed *partially complete chain graph* (PCG) we define in the paper. However, the members diverge about the actual conditional probability distributions for the variables in the common PCG. The combination algorithm we suggest they adopt is one which demands, at least on learning information which is common to the members and which preserves the originally agreed PCG structure, that the pools of *conditional* distributions associated with the PCG are externally Bayesian (EB). We propose a characterization for such conditionally EB (CEB) poolings which is more general and flexible than the characterization proposed by Genest, McConway and Schervish. In particular, such a generalization allows the weights attributed to the joint probability assessments of different individuals in the pool to differ across the distinct components of each joint density. We show that the group's commitment to being CEB on *chain elements* can be accomplished by the group being EB on the whole PCG when the group also agrees to perform the conditional poolings in an ordering compatible with evidence propagation in the graph.

1. Introduction. Bayesian statistical theory is a coherent normative methodology for the individual. However, groups of individuals are left in difficulty since no concept equivalent to the classical notion of objectivity is available to them. Consequently, methods arising either from the axiomatic or the Bayesian modelling approaches [Winkler (1986)] have been developed to attempt to deal with this problem.

We adopt here the *axiomatic approach* to the aggregation of *beliefs* where the group is assumed to choose certain “desirable” and “reasonable” conditions that the aggregating formula should obey and which at the end characterize its form. Early important works in this area are those of Madansky (1964, 1978), who proposed the external Bayesianity property (see Section 2) as a Bayesian prescription for group belief consensus. Later, Bacharach

Received August 1995; revised November 1996.

¹Supported by Grant 200918 \ 92.3 from CNPq, the Brazilian research council.

AMS 1991 *subject classifications*. Primary 62C10; secondary 62H99.

Key words and phrases. Bayesian multivariate decision analysis, expert systems, external Bayesianity, logarithmic opinion pools, chain graphs.

(1972, 1975) introduced the *logarithmic opinion pool* (LogOp) as well as some work on the implication axioms have on pooling rules. Genest (1984) and Genest, McConway and Schervish (1986) characterized the LogOps as the only pools being EB. Also, McConway (1981) showed the linear opinion pool (LinOP) of Stone (1961) as the only type which satisfied an axiom called the marginalization property. The reader should refer to Genest and Zidek (1986) for an excellent survey and annotated bibliography on the axiomatic approaches.

Other important results in the area are the impossibility theorems obtained by Bacharach (1975), Dalkey (1972, 1975) and Lehrer and Wagner (1983), among others. In general, those impossibility results mean that no pooling rule for preferences or beliefs works sensibly well in *all* situations. Combining rules general enough to satisfy certain “desirable” axioms, on the other hand, can fail to be democratic. More closely related to our work, Genest and Wagner (1984) proved that only dictatorships could accommodate both the *likelihood principle* [see (3) in Section 2] and the *independence preservation property* (IPP) [Laddaga (1977)]. Basically, the IPP demands the pooling on joint quantities to equal the product of the poolings marginally in each quantity (in the same spirit of the statistical independence for probability functions), whenever the individuals preserve the quantities’ independence on their subjective assessments.

Since the arrival of these results, there have been two lines of work which attempt to circumvent the impossibility theorems. One line has attempted to weaken group behavioral assumptions [e.g., Pill (1971)] in the hope that the impossibility results would vanish. This has had little success [Winkler (1968)]. The other line, to which this paper is a contribution, has looked at weakening the insistence on universal applicability. Here, our assumption that the group agrees that a specific belief network is valid and is examined only in the light of experiments which do not destroy its structure, prevents some of the impossibility results. In particular, the above-mentioned impossibility of Genest and Wagner (1984) is circumvented [Faria (1996)]. Despite the restrictions imposed on our problem and thus on the pooling operators discussed here, they are of a general enough class to be useful in certain practical situations. For instance, it is easy to devise situations where a common belief net would be an obvious assumption. Moreover, the side condition of the decomposability of the common belief net is not critical since a nondecomposable net can always be made decomposable, as we shall see.

The class of decomposable graphical models introduced by Lauritzen, Speed and Vijayan (1984) is the one which, under certain restrictions on the form of input data, retains its structure (coded in terms of conditional independence statements) after data has been observed. This fact is used extensively to create quick algorithms for calculating posterior distributions in high-dimensional problems [e.g., Dawid (1992), Jensen, Jensen and Dittmer (1994) and Smith and Papamichail (1996)]. However, here we use this same property to define classes of combination rules which are, in a partial sense, externally Bayesian.

In this text we concentrate on generalizing the external Bayesianity property so that on the one hand it is suitable for the analysis of multivariate structures, and on the other it allows the group to learn about the combination's vector of weights. This responds, at least in part, to some of the criticisms [Lindley (1985)] about the inflexibility of the weights on the well-discussed logarithmic pool [Genest and Zidek (1986)] which usually are not allowed to vary with the members' individual expertise. Some examples on how the weights can be updated based on the individuals' relative expertise on causal variables are given in Faria and Smith (1994, 1996) and in Faria (1996). In fact, the weights should be chosen on the basis of the common information that the members of the group have about their own specialities and expertises. A group which cannot or is not prepared to agree about the members' relative expertise may prefer to adopt a class of combination rules which sequentially adapts the weights based on each experts relative predictive performances, as done, for example, by Smith and Makov's (1978) *quasi-Bayes* approach [see, e.g., Faria and Souza (1995) and Faria (1996)]. Other methods on appropriate updating of weights can be found, for example, in Bates and Granger (1969), in Mendel and Sheridan (1987), in Bayarri and DeGroot (1988), in DeGroot and Mortera (1991) and in Cooke (1991). Also, Cooke (1991) comments on the main problems in developing a theory of weights. Although developed for linear pools, there is in principle no reason why sequential methods should not be adapted to be employed for time series in our generalized logarithmic pools where the weights can be functions of past data [see Faria (1996) for comments on this].

The externally Bayesian group has some advantages. For instance, Raiffa (1968) illustrated how the relevance over the order in which the pooling and updating are done can lead to subjects trying to increase their influence on the consensus by insisting that their opinions be computed before the outcome of an experiment is known. If a group is externally Bayesian (EB), then all such argument would be pointless. Also, as Genest (1984) points out, if such a pool can be agreed, then it has great practical advantages. For example, its members need not meet again after data has been observed. However, the main reasons we have chosen the external Bayesianity criterion on which to base this methodology is twofold. First, because *conditional independence* (CI) statements are at the heart of all belief networks, the types of multiplicative pools which result from external Bayesianity tend to preserve the CI structures which relate to *chain graph* (CG) associations, whereas other pools, like the linear opinion pools [McConway (1981)] do not. Second, using a loosened form of the external Bayesianity criterion with a CG structure allows a generalization of the *modified logarithmic opinion pool* (modified LogOp) of Genest, McConway and Schervish (1986) that answers some of the criticisms of its use.

The structure of the paper is as follows. In Section 2 we review the external Bayesianity property and its logarithmic characterization through pooling operators. In Section 3 we define *partially complete chain graphs* (PCG's) and state the conditions under which their structure is preserved

after sampling. A definition of conditional external Bayesianity together with its characterization are proposed in Section 4. In Section 5 we show how CEB poolings in chain elements can characterize an EB pooling on the PCG. The discussion of the results and the plausibility of our assumptions are shown in Section 6. The Appendices contain the proofs of some of the stated theorems.

2. The EB logarithmic opinion pool. Let (Ω, μ) be a measure space. Let Δ be the class of all μ -measurable functions $f: \Omega \rightarrow [0, \infty)$ $f > 0$, μ a.e. such that $\int f d\mu = 1$. A pooling operator $T: \Delta^m \rightarrow \Delta$ is that one which maps a vector of functions (f_1, \dots, f_m) , where $f_i \in \Delta$ (for all $i = 1, \dots, m$), into a single function also in Δ .

Madansky (1964, 1978) characterized the EB pooling operators as those synthesizing the diverging individual opinions f_1, \dots, f_m into a *group opinion* expressed by a single density $\bar{f} = T(f_1, \dots, f_m)$ that must satisfy

$$(1) \quad T\left(\frac{lf_1}{\int lf_1 d\mu}, \dots, \frac{lf_m}{\int lf_m d\mu}\right) = \frac{lT(f_1, \dots, f_m)}{\int lT(f_1, \dots, f_m) d\mu}, \quad \mu\text{-a.e.},$$

where $l: \Omega \rightarrow (0, \infty)$ is the group's common likelihood for (f_1, \dots, f_m) such that $0 < \int lf_i d\mu < \infty$, $i = 1, \dots, m$.

Genest, McConway and Schervish (1986) proved that the following *modified logarithmic opinion pool* is, under certain regularity conditions discussed later, the most general EB logarithmic pool satisfying condition (1) above:

$$(2) \quad T(f_1, \dots, f_m) = \frac{g \prod_{i=1}^m f_i^{w_i}}{\int g \prod_{i=1}^m f_i^{w_i} d\mu}, \quad \mu\text{-a.e.},$$

where $g: \Omega \rightarrow [0, \infty)$ $g > 0$, μ a.e. is an arbitrary bounded function on Ω and w_i , $i = 1, \dots, m$, are arbitrary weights (not necessarily nonnegative) adding up to 1. The w_i 's are experts' opinions weights in the combination and must be suitably chosen to reflect relative expertise. They should possibly be elicited by the experts based on their *common knowledge* set [see, e.g., Geanakoplos (1992)] of their own relative predictive capabilities.

One of the above-mentioned regularity conditions on (Ω, μ) is that for an existing Lebesgue measurable function $P: \Omega \times (0, \infty)^m \rightarrow (0, \infty)$, the pooling operator $T: \Delta^m \rightarrow \Delta$ is such that

$$(3) \quad T(f_1, \dots, f_m)(x) = \frac{P[x, f_1(x), \dots, f_m(x)]}{\int P(\cdot, f_1, \dots, f_m) d\mu}, \quad \mu\text{-a.e.}$$

This condition is called a *likelihood principle* in the sense that it restricts the likelihood of the combined density T at a particular point x in Ω to depend, except for a normalizing constant, on the x and on the individual densities f_i , $i = 1, \dots, m$, assigned to x only through their values at x , and not upon other points and densities of the points which might have occurred but did not. The likelihood principle represented by (3) is a strong and rather arbitrary requirement particularly in the case where T is a multivariate density [e.g., Faria and Smith (1994, 1996)]. In the generalization of EB and

CEB pools we make in Section 4, we require (3) to hold when applied to marginal and conditioned variables related to individual nodes in a graphical representation structure.

Another assumption made here is that the underlying space (Ω, μ) can be partitioned into at least four nonnegligible sets. In this case such a measure space is called *quaternary*. This therefore includes the case where μ is Lebesgue and many (but not all) cases where μ is a counting measure.

McConway (1978) proved that in the case where Ω is countable and μ is a counting-type (purely atomic) measure then the formulas of the type (2) would be the only ones to qualify as EB if (3) holds. In the case where Ω is purely continuous (excluding thus the case where the sample space is countable), Genest (1984) showed that the only nondictatorship externally Bayesian pooling operator satisfying (3) applied to a P that is not indexed by x , that is, $T(f_1, \dots, f_m)(x) = P[f_1(x), \dots, f_m(x)] / \int P(f_1, \dots, f_m) d\mu$, μ -a.e., is Bacharach's (1972) *logarithmic opinion pool* (LogOp):

$$(4) \quad T(f_1, \dots, f_m) = \frac{\prod_{i=1}^m f_i^{w_i}}{\int \prod_{i=1}^m f_i^{w_i} d\mu}, \quad \mu\text{-a.e.},$$

where $w_i \geq 0$, $i = 1, \dots, m$, are *arbitrary constants* such that $\sum_{i=1}^m w_i = 1$.

Unfortunately, as currently developed, EB pools have serious drawbacks. Perhaps the most obvious one is that, as Lindley (1985) points out, it appears perverse that the weights $\underline{w} = (w_1, \dots, w_m)$ must be common knowledge to the group a priori. Surely as evidence appears which sheds light on the relative expertise of its members, the group should agree to adapt its weights to favor the better forecasters [e.g., Faria and Smith (1994, 1996)]. It should be natural that, given the members' commonly held information about their specialities, relative expertise and predictive capabilities, the weights associated with some variables be allowed to depend on the value of other variables in the problem. As we mentioned in the Introduction, our methodology allows for this (see also Section 4). First, we shall introduce the conditions under which the class of the special chain graphs we adopt, the PCG's, have their structures preserved posterior to the input of observed data.

3. The preservation of PCG structures. In this section we define a useful subclass of graphical models, the PCG's (partially complete chain graphs), and an associated family of likelihood functions. Provided the members of the group are able to agree on a causal ordering of the variables and on a dependence structure between them, as embodied in a chain graph (CG), then a class of combination rules much richer than (but containing a special case of) the EB combination rules given in (2) can be obtained. The way this is achieved is by demanding a weakened form of external Bayesianity to hold only when the agreed likelihood does not destroy the conditional independence (CI) structure on the variables represented by the group's agreed PCG.

The properties about the set of uncertain measurements $\underline{X} = (X_1, \dots, X_n)$ stated below are assumed to be common knowledge (CK) to all members of

the group. In fact, there are two basic required conditions about the beliefs of the group's members. The first is that all members agree on the association structure on \underline{X} such that the following holds.

PROPERTY 3.1. (i) The vector \underline{X} can be represented as an ordered list of subvectors $\underline{X}_1, \dots, \underline{X}_n$, where it is CK to all members of a group G that, for $j = 1, \dots, n$, the random vector \underline{X}_j receives a directed association from (or more loosely is *caused by*) $\pi(\underline{X}_j)$, where $\pi(\underline{X}_j)$ is a subvector of $(\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{j-1})$, henceforth called the *group parent set of the chain element* \underline{X}_j .

(ii) It is CK to the group G that \underline{X}_j is conditionally independent of the elements in $(\underline{X}_1, \dots, \underline{X}_{j-1})$ which are not in $\pi(\underline{X}_j)$, given its parent set $\pi(\underline{X}_j)$, that is, in the usual notation of Dawid (1979),

$$\underline{X}_j \perp\!\!\!\perp \{(\underline{X}_1, \dots, \underline{X}_{j-1}) \setminus \pi(\underline{X}_j)\} \mid \pi(\underline{X}_j), \quad j = 1, \dots, n.$$

Assuming that all members have a common dominating measure we can then assert that the i th member of the group can write his joint density over \underline{X} , $f_i(\underline{x})$, in the form

$$f_i(\underline{x}) = \prod_{j=1}^n f_{ij}[\underline{x}_j \mid \pi(\underline{x}_j)], \quad i = 1, \dots, m,$$

where \underline{x}_j and $\pi(\underline{x}_j)$ are defined above for $j = 2, \dots, n$ and $\pi(\underline{x}_1) = \emptyset$. Henceforth we shall impose the usual positivity condition that $f_{ij}[\underline{x}_j \mid \pi(\underline{x}_j)] > 0$ for each value of $\pi(\underline{x}_j)$ at all values of \underline{x}_j in this space, which is common to all members of the group.

The association structure described by (i) and (ii) above defines a class of multivariate structures which are a subclass of graphical models called *chain graph* (CG) models. Briefly, a CG is a graph which does not contain any directed cycles and whose set of nodes (vertices) can be partitioned into numbered subvectors. These subvectors form a *dependence chain* such that all associations between nodes within the same subvector are undirected (forming the *chain elements* of the CG) and all associations between components of distinct subvectors are directed. A directed edge from the component of a lower labelled vector points to a component of a higher labelled vector. The literature on these models is extensive [e.g., Wermuth and Lauritzen (1990), Frydenberg (1990) and Lauritzen (1996)].

The PCG's are defined as follows [see Faria and Smith (1994) or Faria (1996) for an example].

DEFINITION 3.2 (PCG). A chain graph $\mathcal{G}^p(\underline{X})$ with nodes labelled by its chain elements $\underline{X}_1, \dots, \underline{X}_n$ is called a *partially complete chain graph* if, for each $j = 1, \dots, n$, all the components of the subvector (chain element) \underline{X}_j of $\mathcal{G}^p(\underline{X})$ are connected together to form a *complete* undirected subgraph. A directed edge connects $\underline{X} \in \underline{X}_i$ to $\underline{Y} \in \underline{X}_j$ if and only if \underline{X}_i is a subvector of $\pi(\underline{X}_j)$, where $\pi(\underline{X}_j)$ is the parent set of \underline{X}_j .

The PCG's are ones which, by judicious changes of definition, can be represented by influence diagrams (IDs) on vectors of variables [e.g., Smith (1989) and Queen and Smith (1993)]. The ID *induced* by the PCG defined above consists of the statements of Property 3.1 together with a directed graph whose n nodes are labelled $\underline{X}_1, \dots, \underline{X}_n$ and where \underline{X}_i is connected by an edge to \underline{X}_j if and only if \underline{X}_i is a subvector of $\pi(\underline{X}_j)$.

We shall call a PCG *decomposable*, $\mathcal{G}^{pd}(\underline{X})$, if the ID induced by $\mathcal{G}^p(\underline{X})$ is decomposable, that is, when for some k ($k = 1, \dots, n$), $\underline{X}_i, \underline{X}_j$ are both subvectors of $\pi(\underline{X}_k)$, $i \neq j$, then either \underline{X}_i is a subvector of $\pi(\underline{X}_j)$ or \underline{X}_j is a subvector of $\pi(\underline{X}_i)$.

Decomposable models are very common in statistics. They include the cases when (a) all variables are independent of one another, (b) they are all dependent on one another and (c) they form a Markov chain. They also include most hierarchical models like the dynamic linear model [West and Harrison (1989)]. If a model is not originally decomposable it can be made so by ignoring certain stated conditional independencies [see Lauritzen and Spiegelhalter (1988) for an example of how to do this].

The *moral graph* obtained from a CG $\mathcal{G}(\underline{X})$ is the undirected graph $\mathcal{G}^m(\underline{X})$ with the same set of nodes \underline{X} but with any pair of nodes being joined together by an undirected edge in \mathcal{G}^m if and only if (i) they were already joined in \mathcal{G} (by either a directed or an undirected edge) or (ii) they both were parents of nodes in the same chain element. Thus, the moral graph of a directed acyclic graph is built up by joining with undirected edges all nodes that have a common *child* and by replacing all directed edges by undirected ones [see, e.g., Lauritzen and Spiegelhalter (1988)].

Since the components of the chain elements \underline{X}_j of a PCG $\mathcal{G}^p(\underline{X})$ form a complete undirected graph, any new information about a set of such components is informative about all the other components of that chain element and no CI assumption is destroyed within that chain element.

In order to generalize external Bayesianity, it is also necessary to introduce a second condition, which will act on the class of likelihoods for models defined on a PCG $\mathcal{G}^p(\underline{X})$. In a general setting, these likelihoods, which we shall call *cutting*, are those which can be only informative about one chain element \underline{X}_j and/or its parents $\pi(\underline{X}_j)$ in the group's common PCG.

DEFINITION 3.3 (Cutting likelihood). Say that $l(\underline{x} | \underline{z})$ is in the class of *cutting likelihoods* related to a PCG $\mathcal{G}^p(\underline{X} | \underline{Z})$, henceforth denoted by $\mathcal{L}(\mathcal{G}^p)$, if it is a likelihood function which could have resulted from a sample $\underline{Z} = (\underline{Z}_1, \dots, \underline{Z}_n)$ whose density $g(\underline{z} | \underline{x})$ can be written in the following form:

$$(5) \quad g(\underline{z} | \underline{x}) = g_1(\underline{z}_1 | \underline{x}_1) g_2[\underline{z}_2 | \underline{x}_2, \pi(\underline{x}_2), \underline{z}_1] \\ \times \dots g_n[\underline{z}_n | \underline{x}_n, \pi(\underline{x}_n), \underline{z}^{n-1}],$$

where $\underline{x} = (\underline{x}_1, \dots, \underline{x}_n)$ are values of the components of \underline{X} in \mathcal{G}^p , $\pi(\underline{x}_j)$ are fixed values of the parents of \underline{X}_j in $\mathcal{G}^p(\underline{X})$ and $\underline{z}^k = (\underline{z}_1, \underline{z}_2, \dots, \underline{z}_k)$, for $k \geq 1$, are the observed values of $\underline{Z}^k = (\underline{Z}_1, \dots, \underline{Z}_k)$.

If \mathcal{G}^p is complete, then $L(\mathcal{G}^p)$ is the class of all likelihoods. So we only constrain our class of likelihoods when there is some substantive agreement between the members about some lack of association between certain sets of variables given another. In the case when $\mathcal{G}^p(\underline{X})$ is not complete, there is an ordering of \underline{X} to $(\underline{X}_1, \dots, \underline{X}_n)$ such that \underline{X}_i can be thought (loosely) of as being caused by $\pi(\underline{X}_i)$, a subvector of $(\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{i-1})$, $i = 2, \dots, n$. When $l(\underline{x} | \underline{z}) \in \mathcal{L}(\mathcal{G}^p(\underline{X} | \underline{Z}))$ we assume that we have taken an observation \underline{Z} which, possibly after some transformation, can be represented as $(\underline{Z}_1, \dots, \underline{Z}_n)$, where \underline{Z}_i is dependent on \underline{X} only through \underline{X}_i and the values of its direct causes $\pi(\underline{X}_i)$. The most extreme case is when $\underline{X}_1, \dots, \underline{X}_n$ are independent. In this case a cutting likelihood is one which separates in each of these variables. In the case of noncomplete PCG's, the ID induced from $\mathcal{G}^p(\underline{X})$ must be decomposable otherwise original relevances may not be valid anymore after sampling, as we shall see.

Now, the second mentioned condition the group is required to obey regarding the likelihoods obtained from sampling over a decomposable PCG is the following:

PROPERTY 3.4. External Bayesianity holds only with respect to incoming information \underline{Z}_j , about the chain elements X_j of a decomposable PCG $\mathcal{G}^{pd}(\underline{X})$, for which the value of the ancestral set $\alpha(\underline{z}_j) = (\underline{z}_1, \dots, \underline{z}_{j-1})$ is already known, and whose likelihood $l_j[\underline{x}_j | \underline{z}^j, \pi(\underline{x}_j)]$, $j = 1, \dots, n$, is a component of a cutting likelihood $l(\underline{x} | \underline{z})$ related to $\mathcal{G}^{pd}(\underline{X})$.

In particular, the form (5) prevents a variable \underline{X}_k in a decomposable PCG, \mathcal{G}^{pd} , not belonging to the parent set $\pi(\underline{X}_j)$ of \underline{X}_j [which is associated with the index of the product component g_j in (5)], to condition the observation \underline{z}_j in $g_j[\underline{z}_j | \underline{x}_j, \pi(\underline{x}_j), \underline{z}^{j-1}]$. In other words, no direct associations between $\underline{X}_k \notin \pi(\underline{X}_j)$ and \underline{Z}_j are allowed in $\mathcal{G}^{pd}(\underline{X} | \underline{Z})$. This, in its turn, avoids the introduction of any new association in the moral graph of $\mathcal{G}^{pd}(\underline{X} | \underline{Z})$ not originally present in $\mathcal{G}^{pd}(\underline{X})$. Therefore, no CI is lost in \mathcal{G}^{pd} according to Pearl's (1988) d-separation theorem.

The class of cutting likelihoods is a very natural one to consider in the context of PCG's, for CG's \mathcal{G}^p whose induced ID is decomposable, \mathcal{G}^{pd} . This is because, provided Property 3.4 is satisfied by the group, the class is determined by those data sets which, for each member, are guaranteed to preserve the CI structure implicit in \mathcal{G}^{pd} after data assimilation. Thus it is simply information which does not destroy the association structure agreed by members of the group. Now we present the formal statement and proof of this result.

THEOREM 3.5. (a) *If Property 3.4 is satisfied by the group G for likelihood functions $l(\underline{x} | \underline{z})$ related to the variables of a PCG whose induced ID is decomposable, $\mathcal{G}^{pd}(\underline{X})$, then for each member of G , $\mathcal{G}^{pd}(\underline{X} | \underline{Z})$ is a CG of that member's joint density of $\underline{X} | \underline{Z}$.*

(b) *If Property 3.4 is not satisfied, then, for some member of G , $\mathcal{G}^{pd}(\underline{X} | \underline{Z})$ may not be the CG of $\underline{X} | \underline{Z}$.*

PROOF. (a) First we draw the ID $\mathcal{S}(\underline{X}, \underline{Z})$ whose nodes are $(\underline{Z}_1, \dots, \underline{Z}_n, \underline{X}_1, \dots, \underline{X}_n)$. The ID of $\underline{X}_1, \dots, \underline{X}_n, \mathcal{S}(\underline{X})$, is the one induced by $\mathcal{G}^{pd}(\underline{X})$. A node \underline{Y} is a parent of \underline{Z}_j , $j = 1, \dots, n$, if either (i) $\underline{Y} = \underline{X}_j$ or $\underline{Y} \in \pi(\underline{X}_j)$ or (ii) $\underline{Y} = \underline{Z}_k$, $k = 1, \dots, j - 1$.

Now use Pearl's (1988) d-separation theorem. New edges inducing the marriage of parents of \underline{Z}_j , $j = 1, \dots, n$, can only occur between a Z node and an X node. This is because $\mathcal{G}^{pd}(\underline{X}, \underline{Z})$ has an induced decomposable ID within $\mathcal{S}(\underline{X}, \underline{Z})$, and the subgraph on subvectors of $[\pi(\underline{X}_j), \underline{X}_j]$ is complete. It follows that all new paths in $\mathcal{S}(\underline{X}, \underline{Z})$ between nodes on the subgraph $\mathcal{S}(\underline{X})$ of $\mathcal{S}(\underline{X}, \underline{Z})$ induced by the marriage of parents are blocked by $\{\underline{Z}_1, \dots, \underline{Z}_n\}$. Also, there will be no unblocked new path linking X nodes which were not linked in the original association structure. This is because the value of the ancestral set $\alpha(\underline{Z}_j)$ is assumed to be already known when \underline{Z}_j is observed. Therefore if a CI statement is implied in $\mathcal{S}(\underline{X})$ on \underline{X} , then it is also implied on $\mathcal{S}(\underline{X}, \underline{Z})$.

(b) If $l(\underline{x} | \underline{z}) \notin \mathcal{L}(\mathcal{G}^{pd})$, then for some index j , $j = 1, \dots, n$, there exists a \underline{Z}_j such that the parents of \underline{Z}_j will have an edge induced to \underline{X}_j after moralization. Marrying the parents of the conditioning \underline{Z}_j will now produce an unblocked path between another node \underline{X}_k , $k = 1, \dots, j$, not connected to \underline{X}_j in $\mathcal{S}(\underline{X})$, and \underline{X}_j . So, whereas all members agreed a priori that $\underline{X} \perp\!\!\!\perp \underline{X}_k | \pi(\underline{X}_j)$, after observing \underline{Z}_j this can no longer be deduced. \square

4. Characterization of conditionally externally Bayesian pooling operators. Suppose that the m members of a group have agreed on the structure of a decomposable PCG $\mathcal{G}^{pd}(\underline{X})$ relating n random vectors $\underline{X} = (\underline{X}_1, \dots, \underline{X}_n)$ in a certain problem. Let \mathbf{A}_j be the event that the parent nodes of \underline{X}_j , $\pi(\underline{X}_j)$, have fixed values \underline{X}_j , that is, $\mathbf{A}_j = \{\pi(\underline{X}_j) = \pi(\underline{x}_j)\}$ for $j = 1, \dots, n$. Despite believing the common PCG, \mathcal{G}^{pd} , each member has its own particular opinion about the parameters of his conditional densities $f_{ij}(\underline{X}_j | \mathbf{A}_j)$ associated with the graph structure. For technical reasons we shall assume that $f_{ij} > 0$ for $i = 1, \dots, m$ and $j = 1, \dots, n$. The members agree to follow the external Bayesianity axiom, satisfying thus (1) applied to for sets of conditional densities. Hence, according to the probability breakdown in CG's, external Bayesianity on $\underline{X}_j | \mathbf{A}_j$ is required for all possible values of the \underline{X}_j , $j = 1, \dots, n$. Unlike in (1), the likelihood over which we demand external Bayesianity to hold for these conditional densities is restricted to the family $\mathcal{L}(\mathcal{G}^{pd})$ of cutting likelihoods defined in Section 3. Thus we demand that Property 3.4 is satisfied. Particularly, this means that external Bayesianity is required *only* for new information that might come from a designed experiment whose design points are the parents or causes of that variable.

We can now define the conditional external Bayesianity property which will characterize conditionally externally Bayesian (CEB) pooling operators representing combined probability density functions associated with a PCG.

DEFINITION 4.1 (Conditional external Bayesianity). Say a group G obeys the *conditional external Bayesianity* property if the joint density $f(\underline{x})$ of the

variables in its common decomposable PCG $\mathcal{G}^{pd}(\underline{X})$ is combined in the following way. For each component \underline{X}_j of \underline{X} and each set \mathbf{A}_j of possible values of the parents $\pi(\underline{X}_j)$ of \underline{X}_j , each of the conditional densities $f_j(\underline{x}_j | \mathbf{A}_j)$ is pooled to preserve the external Bayesianity property ($j = 1, \dots, n$) with respect to the component $l_j(\underline{x}_j | \mathbf{A}_j, \underline{z}^j)$ of the common cutting likelihood $l(\underline{x} | \underline{z})$ associated with $\mathcal{G}^{pd}(\underline{X})$.

In terms of obtaining a characterization of conditional external Bayesianity through a class of pooling operators, we propose, in line with Madansky (1964, 1978) and Genest McConway and Schervish (1986), that a CEB pooling operator T_j associated with a PCG $\mathcal{G}^p(\underline{X})$ is one which is EB on appropriate conditional distributions and only for data that respect an ordering of that conditioning, $\underline{X}_j | \mathbf{A}_j$, $j = 1, \dots, n$, implicit in that chain when Property 3.4 is satisfied.

Define a measure space $[\Omega_j(\mathbf{A}_j), \mu_j^*(\mathbf{A}_j)]$, hereafter denoted (Ω_j, μ_j^*) , with Ω_j being the product space of spaces related to components of \underline{X}_j and μ_j^* being the product reference measure associated with the $r(j)$ -dimensional vector \underline{X}_j in Ω_j [see, e.g., Rudin (1986)]. Let $T_j: \Delta_j^m \rightarrow \Delta_j$ be a CEB pooling operator, where Δ_j is the class of all μ_j^* -measurable functions $f_{ij}: \Omega_j \rightarrow [0, \infty)$ with $f_{ij} > 0$ (μ_j^* -a.e.) such that $\int \dots \int f_{ij} d\mu_{1j} \dots d\mu_{r(j)j} = 1$ for all i, j and $r(j)$. The μ_{kj} 's are measures associated with components of \underline{X}_j . If such a pooling operator satisfies the following multivariate conditional version of (3):

$$(6) \quad T_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \mathbf{A}_j) = \frac{P_j[\underline{x}_j | \mathbf{A}_j, f_{1j}(\underline{x}_j | \mathbf{A}_j), \dots, f_{mj}(\underline{x}_j | \mathbf{A}_j)]}{\int \dots \int P_j[\cdot, f_{1j}, \dots, f_{mj}] d\mu_{1j} \dots d\mu_{r(j)j}}, \quad \mu_j^*\text{-a.e.},$$

for each $j = 1, \dots, n$, where $P_j: \Omega_j \times (0, \infty)^m \rightarrow (0, \infty)$ is some arbitrary Lebesgue measurable function; then $\underline{T} = (T_1, \dots, T_n)$ is said to satisfy the PCG \mathcal{G}^{pd} likelihood principle. This condition has the same interpretation as condition (3) but for \underline{X}_j conditioned on \mathbf{A}_j ; that is, except for a normalization factor which does not depend on \underline{X}_j , the density of the consensus at $\underline{X}_j | \mathbf{A}_j$ is required to depend only on X_j and its fixed parents $\pi(\underline{X}_j)$ as well as on the individual densities at the actual value of the unseen quantities given its parents, but not upon the densities of the values which might have obtained but did not.

Similar to Madansky's condition (1) for EB pooling operators, the CEB pooling operator T_j is required to satisfy the following condition [see Faria and Smith (1994) and Faria (1996)]:

$$(7) \quad T_j \left[\frac{l_j f_{1j}}{\int \dots \int l_j f_{1j} d\mu_{1j} \dots d\mu_{r(j)j}}, \dots, \frac{l_j f_{mj}}{\int \dots \int l_j f_{mj} d\mu_{1j} \dots d\mu_{r(j)j}} \right] = \frac{l_j T_j(f_{1j}, \dots, f_{mj})}{\int \dots \int l_j T_j(f_{1j}, \dots, f_{mj}) d\mu_{1j} \dots d\mu_{r(j)j}}$$

with μ_j^* -a.e. for $j = 1, \dots, n$ where the likelihood l_j is a component of $l \in \mathcal{L}(\mathcal{G}^{pd})$.

Assuming the underlying measure space (Ω_j, μ_j^*) of each vector \underline{X}_j in $\mathcal{E}^{pd}(\underline{X})$ can be partitioned into at least four nonnegligible sets (that includes the continuous and most of the countable cases), it is straightforward to extend the characterization theorem of Genest, McConway and Schervish (1986) for such an operator in the following way.

THEOREM 4.2 (Conditional modified LogOp). *Let (Ω_j, μ_j^*) be the quaternary measure space. Let $\tilde{f}_j(\underline{x}_j | \mathbf{A}_j): \Delta_j^m \rightarrow \Delta_j$ be a CEB pooling operator representing the m individuals combined conditional density for the $r(j)$ -dimensional random vector $\underline{X}_j, j = 1, \dots, n$, given its parents in a PCG $\mathcal{E}^p(\underline{X})$. If*

$$\tilde{f}_j(\underline{x}_j | \mathbf{A}_j) = T_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \mathbf{A}_j),$$

$j = 1, \dots, n$, where for all $f_{ij} (i = 1, \dots, m)$ in Δ_j^m and for an existing $\mu_j^* \times$ Lebesgue measurable function $P_j: \Omega_j \times (0, \infty)^m \rightarrow (0, \infty), T_j: \Delta_j^m \rightarrow \Delta_j$ satisfies (6), then \tilde{f}_j takes the form

$$(8) \quad \tilde{f}_j(\underline{x}_j | \mathbf{A}_j) = \frac{p_j \prod_{i=1}^m [f_{ij}(\underline{x}_j | \mathbf{A}_j)]^{w_{ij}(\mathbf{A}_j)}}{\int \cdots \int p_j \prod_{i=1}^m [f_{ij}(\underline{x}_j | \mathbf{A}_j)]^{w_{ij}(\mathbf{A}_j)} d\mu_{1j} \cdots d\mu_{r(j)j}}, \quad \mu_j^* \text{-a.e.,}$$

where $p_j: \Omega_j \rightarrow [0, \infty) P_j > 0, \mu$ a.e. are essentially bounded functions and $w_{ij}(\mathbf{A}_j)$ are weights such that $\sum_{i=1}^m w_{ij}(\mathbf{A}_j) = 1$ holds for each index $j = 1, \dots, n$, and \mathbf{A}_j are the variables whose values are commonly known by the group when the combination rule is applied. Furthermore, the weights are nonnegative unless Ω_j is finite or there does not exist a countably infinite partition of (Ω_j, μ_j^*) into nonnegligible sets.

The proof of this theorem is straightforward by a slight adaptation of Theorem 4.4 in Genest, McConway and Schervish (1986) individually for each node in \mathcal{E}^p associated with the underlying measure space (Ω_j, μ_j^*) , but conditioned on $\underline{X}_j | \mathbf{A}_j, j = 1, \dots, n$. The proof together with some intermediate results is given in Appendix A.

The $w_{ij}(\mathbf{A}_j)$ are the weights that should be a measure of i 's expertise associated with the vector \underline{X}_j . They can be possibly a function of other components in \underline{X}_j .

It is rather difficult to give an interpretation to p_j in the context of our group decision problem. However, it is reasonable in the majority of problems to require that T_j preserves the group's unanimity. This leads us to setting p_j equal 1. Note that in this case condition (6) can be restricted to

$$T_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \mathbf{A}_j) = \frac{P_j[\mathbf{A}_j, f_{1j}(\underline{x}_j | \mathbf{A}_j), \dots, f_{mj}(\underline{x}_j | \mathbf{A}_j)]}{\int \cdots \int P_j[\cdot, f_{1j}, \dots, f_{mj}] d\mu_{1j} \cdots d\mu_{r(j)j}}, \quad \mu_j^* \text{-a.e.,}$$

thus not allowing T_j to depend on \underline{x}_j directly.

The following corollary can be stated, with the proof being easily obtained from Theorem 4.2.

COROLLARY 4.3 (Conditional LogOp). *Let (Ω_j, μ_j^*) be a quaternary measure space and let $T_j: \Delta_j^m \rightarrow \Delta_j$ be a CEB pooling operator which preserves unanimity. If there exists a $\mu_j^* \times$ Lebesgue measurable function $P_j: \Omega_j \times (0, \infty)^m \rightarrow (0, \infty)$ such that T_j satisfies (6) for all vectors of conditional opinions $(f_{1j}, \dots, f_{mj}) \in \Delta_j^m$, then T_j is a logarithmic opinion pool, that is,*

$$(9) \quad T_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \mathbf{A}_j) = \frac{\prod_{i=1}^m [f_{ij}(\underline{x}_j | \mathbf{A}_j)]^{w_{ij}(\mathbf{A}_j)}}{\int \cdots \int \prod_{i=1}^m [f_{ij}(\underline{x}_j | \mathbf{A}_j)]^{w_{ij}(\mathbf{A}_j)} d\mu_{1j} \cdots d\mu_{r(j)j}}, \quad \mu_j^*\text{-a.e.},$$

for some arbitrary weights w_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, possibly functions of \mathbf{A}_j (the commonly known past when the densities of \underline{X}_j are combined) adding up to 1. Moreover, the weights $w_{ij}(\mathbf{A}_j)$ are nonnegative unless Ω_j is finite or there does not exist a countable partition of (Ω_j, μ_j^*) into nonnegligible sets.

5. CEB poolings appearing EB. We have already shown in Section 3 that in order to ensure that the CEB rules are well defined, the common PCG $\mathcal{G}^p(\underline{X})$ must be decomposable. Note that the CEB rules are *not* based on pooling operators since their arguments are not necessarily just the values of the joint densities in those pools.

The question now is, when $l(\underline{x} | \underline{z}) \in \mathcal{L}(\mathcal{G}^{pd})$, in what sense, if any, are the CEB poolings on chain elements EB on the whole PCG \mathcal{G}^{pd} ?

Certainly, when data \underline{Z} about \underline{X} in $\mathcal{G}^{pd}(\underline{X})$ is observed, that evidence must be propagated through the PCG. Therefore, all the conditional poolings $T_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \underline{x}^{(n-1)})$ on the chain elements \underline{X}_j and its predecessors on $\mathcal{G}^{pd}(\underline{X})$, that is, $\underline{X}_1, \dots, \underline{X}_{j-1}$, must be updated to (omitting the members's densities) $T_j[\underline{x}_j | \underline{x}^{(j-1)}, \underline{z}]$ for $j = 1, \dots, n$.

Suppose we demand that the group agrees to update those conditional densities in a backwards sequence. Thus assume that the group agrees to update $\underline{X}_n | \underline{X}^{(n-1)}$ first, $\underline{X}_{n-1} | \underline{X}^{(n-2)}$ second and so on to \underline{X}_1 .

Let \tilde{f} be the group's combined joint density which takes the individual posterior densities on $\underline{X}_n | [\underline{X}^{(n-1)}, \underline{Z}]$, pools them, uses the derived (agreed) density of $\underline{Z} | \underline{X}^{(n-1)}$ to obtain individual densities of $\underline{X}_{n-1} | [\underline{X}^{(n-2)}, \underline{Z}]$, pool these and so on down to the density of \underline{X}_1 . Also, let \tilde{f} be the group's combined joint density which pools the individual prior densities of $\underline{X}_n | \underline{X}^{(n-1)}$, forms a group's posterior density of $\underline{X}_n | [\underline{X}^{(n-1)}, \underline{Z}]$ and a density of $\underline{Z} | \underline{X}^{(n-1)}$, uses this agreed density of $\underline{Z} | \underline{X}^{(n-1)}$ and the pool of the prior densities on $\underline{X}_{n-1} | \underline{X}^{(n-2)}$ to obtain the posterior density $\underline{X}_{n-1} | [\underline{X}^{(n-2)}, \underline{Z}]$ and so on down to the pooling of \underline{X}_1 . Thus, the question is when does $\tilde{f}(\underline{x} | \underline{z}) = \tilde{f}(\underline{x} | \underline{z})$? The answer is provided by the following theorem.

THEOREM 5.1. *Suppose that Property 3.4 is satisfied by the group G for a PCG $\mathcal{G}^{pd}(\underline{X})$. Also assume that the vector of weights \underline{w}_j of the conditional LogOps used to combine the beliefs of the members of G on chain elements \underline{X}_j*

of $\mathcal{G}^{pd}(\underline{X})$ is a function only of variables in $\pi(\underline{x}_j)$ for all $j = 1, \dots, n$. Then, for the whole graph $\mathcal{G}^{pd}(\underline{X} | \underline{Z})$,

$$\tilde{f}(\underline{x} | \underline{z}) = \bar{f}(\underline{x} | \underline{z}),$$

where the conditional LogOps components of \tilde{f} or \bar{f} are backwards sequentially updated.

See Appendix B for the proof of this theorem.

Note that the above result guarantees that the original PCG structure is preserved after new information is incorporated into the model and that there is agreement on how the graph is updated, although the updating is only strictly EB when the corresponding graph is completely disconnected.

6. Discussion. Complete CG's are always decomposable and make no statements about CI being probabilistically valid for all situations. Therefore, for complete CG's, Theorem 5.1 implies that all CEB rules are *sequentially* EB (in the sense defined in the previous section) to general likelihoods. Each ordering of variables in a complete induced ID gives a different class of sequentially EB pool, so with n variables there are $n!$ different classes defined by different CG's. Consequently, the collection of such sequentially EB pools is extremely rich, a fact obscured by the insistence that a pooling should be an operator on a joint density [see Faria and Smith (1996), Example 2.2]. On the other hand, for CG's representing unconditional independence between all the variables, Theorem 5.1 implies that the joint pooling on the whole CG is EB to cutting likelihoods. In this case, the agreement on the sequential backwards performing of conditional pools need not be made.

The CEB pools are not formed as pooling operators on joint densities on all the variables of a system. They act as components of pooling operators on conditional densities. The argument that a pooling should be a pooling operator on a joint density appears to us to be weak, since relative expertise on particular variables should be allowed.

This multiplicity in the complete case is, in one sense, a problem, since we need to choose which CEB pool to use. But this will largely be determined by the time or causal order in which the random variables are observed. We need this ordering to fix the weights w_j associated with the j th variables \underline{X}_j since w_j is allowed only to depend on the parents of \underline{X}_j . In the complete case, $\pi(\underline{X}_j) = \{\underline{X}_1, \dots, \underline{X}_{j-1}\}$, and, for all possible pools to operate in the class of CEB pools associated with this graph, $\underline{X}_1, \dots, \underline{X}_{j-1}$ will not need to be known before the pooling takes place [Faria (1996) and Faria and Smith (1996)].

When the induced ID is not decomposable, assimilation of data tends to destroy that agreed structure. It is natural therefore to work only with information in the ID which is not destroyed by assimilation of new information.

On the other hand, an agreed nondecomposable ID \mathcal{I} can always be made decomposable by marrying parents and adding directed edges until it is decomposable [Lauritzen and Spiegelhalter (1988)]. By using this derived ID

\mathcal{S}^d as a basis for the CEB pools instead of \mathcal{S} , some information might be lost because there may be CI statements supplied by \mathcal{S} but not by \mathcal{S}^d . Rather we claim that it is only sensible to formulate explicitly the pooling on those agreed conditional independencies which can reasonably be assumed to be preserved after simple types of sampling. Otherwise, the combination obtained a posteriori would not even be defined [see Faria and Smith (1996), Example 2.1].

One legitimate question that might be asked is: why use CG's whose induced ID's are not complete? Because an ID is always acyclic there will always be a complete dimensional graph which has it as a subgraph and so is a valid (ID) description of the problem. Furthermore, the CEB pools which are associated with complete graphs have the advantage that they are sequentially EB with respect to all likelihoods. On the other hand, CEB pools related to incomplete graphs are only sequentially EB to data which gives rise to a certain structure of likelihood. There are four answers to this question:

1. *Simplicity.* If the type of information you expect to receive will automatically preserve CI structure, it seems perverse to demand methods of combining densities which exhibit individual's dependence structures they will never believe.
2. *Preservation of symmetry.* Suppose two random variables X_i, X_j in the random vector \underline{X} are agreed to exhibit CI and are symmetric in the sense that there is no clear order of causality or association between them. Thus, to introduce such an association into the pooling algorithms seems to be artificial and undesirable.
3. *Fixing a frame.* In a given problem, experts will agree on a set of random vectors $\underline{X}_1, \dots, \underline{X}_n$ on which they will pool their opinions. However, in most circumstances, each expert will have beliefs about other variables \underline{X}^* , agreed as independent of \underline{X} . Implicitly in any pooling, they will ignore the disparity between their beliefs about \underline{X}^* . Similarly, at a future time, if asked to combine their beliefs about \underline{X}^* , they will choose to ignore the disparity in their beliefs about \underline{X} . But to do this implies the use of a CEB rule which explicitly demands in its associated CG that $\underline{X} \perp\!\!\!\perp \underline{X}^*$. Thus, if we do not allow incomplete cases, in different and independent forecasting problems about \underline{X} and \underline{X}^* , then we would need to prioritize \underline{X} and \underline{X}^* .
4. *External Bayesianity.* The sparser the CG, the more informative it is in terms of CI statements and the closer to being EB the group is in the joint combination.

APPENDIX A

Proof of Theorem 4.2. The proof of Theorem 4.2 needs intermediate results that we state here as other theorems. Basically, the proof itself is split up to include all possible configurations of the underlying measure spaces, that is, the cases in which (Ω_j, μ_j^*) does not contain any atoms, or is purely atomic, or contains atoms but is not purely atomic. Some of these intermedi-

ate theorems themselves need other results and concepts. The complete proof together with the proofs of all related lemmas and theorems can be found in Faria and Smith (1994) or in Faria (1996).

We begin Section A.1 with two characterization theorems of the *conditional modified LogOp*. Theorem A.1 covers the case in which (Ω_j, μ_j^*) is not purely atomic and condition (6) holds for the functions $P_j, j = 1, \dots, n$. Theorem A.2 characterizes (8) for the case in which (Ω_j, μ_j^*) contains at least three atoms but is not purely atomic. Those two theorems are then used in Section A.2 to prove Theorem 4.2.

A.1. *Some characterization theorems for the conditional modified LogOp.*

THEOREM A.1. *Assume that (Ω_j, μ_j^*) is not purely atomic and that N is the complement set of the atoms. Let $\tilde{f}_j: \Delta^m \rightarrow \Delta, j = 1, \dots, n$, be externally Bayesian. Assume that (6) holds for $\mu_j^* \times$ Lebesgue measurable functions $P_j: \Omega_j \times (0, \infty)^m \rightarrow (0, \infty), j = 1, \dots, n$. Then*

$$(10) \quad \tilde{f}_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \mathbf{A}_j) = \frac{p_j(\underline{x}_j | \mathbf{A}_j) \prod_{i=1}^m [f_{ij}(\underline{x}_j | \mathbf{A}_j)]^{w_{ij}(\mathbf{A}_j)}}{\int \dots \int P_j(\cdot, f_{1j}, \dots, f_{mj}) d\mu_{1j} \dots d\mu_{r(j)j}},$$

μ_j^* -a.e. on N ,

where $w_{ij}(\mathbf{A}_j) \geq 0$ and $\sum_{i=1}^m w_{ij}(\mathbf{A}_j) = 1, j = 1, \dots, n$.

See Faria and Smith (1994) or Faria (1996) for a complete proof of this theorem which is a straightforward generalization of Lemma 4.3 in Genest, McConway and Schervish (1986).

THEOREM A.2. *Let (Ω_j, μ_j^*) be a quaternary measure space that contains at least two atoms, and let $\tilde{f}_j: \Delta^m \rightarrow \Delta$ be a CEB pooling operator. Suppose there exist $\mu_j^* \times$ Lebesgue measurable functions $P_j: \Omega_j \times (0, \infty)^m \rightarrow (0, \infty)$ such that (6) holds for all conditional densities $f_{ij} \in \Delta, i = 1, \dots, m$ and $j = 1, \dots, n$, associated with a PCG \mathcal{G}^p . Then there exist, for each vector of atoms $\underline{X}_j | \mathbf{A}_j$ in $\Omega_j, j = 1, \dots, n$, constant terms $v_{1j}(\mathbf{A}_j), \dots, v_{mj}(\mathbf{A}_j)$ such that*

$$(11) \quad \tilde{f}_j(f_{1j}, \dots, f_{mj})(\underline{x}_j | \mathbf{A}_j) = \frac{p_j(\underline{x}_j | \mathbf{A}_j) \prod_{i=1}^m [f_{ij}(\underline{x}_j | \mathbf{A}_j)]^{v_{ij}(\mathbf{A}_j)}}{\int \dots \int p_j \prod_{i=1}^m f_{ij}^{v_{ij}(\mathbf{A}_j)} d\mu_{1j} \dots d\mu_{r(j)j}}.$$

PROOF. Lemma B.8 in Faria and Smith (1994) [a generalization of Lemma 4.2 in Genest, McConway and Schervish (1986)] proves that, for the hypotheses above, there exist, for each pair of vectors of atoms $(\underline{X}_j | \mathbf{A}_j, \underline{Y}_j | \mathbf{D}_j)$ in Ω_j^2 , constant terms $v_{1j}(\mathbf{A}_j, \mathbf{D}_j), \dots, v_{mj}(\mathbf{A}_j, \mathbf{D}_j)$ such that, for all functions $g_{1j}, \dots, g_{mj} > 0$ in Δ , we have for each j that

$$Q_j(\underline{x}_j | \mathbf{A}_j, \underline{y}_j | \mathbf{D}_j)(g_{1j}, \dots, g_{mj}) = Q_j(\underline{x}_j | \mathbf{A}_j, \underline{y}_j | \mathbf{D}_j)(\mathbf{1}) \prod_{i=1}^m g_{ij}^{v_{ij}(\mathbf{A}_j, \mathbf{D}_j)},$$

where $Q_j: (0, \infty)^m \rightarrow (0, \infty)$ are Lebesgue measurable functions and $\underline{1} = (1, \dots, 1)$ is m -dimensional. So, if for all j we fix the vector of atoms $y_j | \mathbf{D}_j \in \Omega_j$ and choose $0 < \varepsilon_j < 1/\mu_j^*(y_j | \mathbf{D}_j)$, then we can define for all atoms $\underline{x}_j | \mathbf{A}_j \in \Omega_j$ the functions

$$P_j(\underline{x}_j | \mathbf{A}_j) = Q(\underline{x}_j | \mathbf{A}_j, y_j | \mathbf{D}_j)(\underline{1})P_j[y_j | \mathbf{D}_j, \varepsilon_j] \varepsilon_j^{-v_j(\mathbf{A}_j, \mathbf{D}_j)},$$

where $\varepsilon_j = (\varepsilon_{j1}, \dots, \varepsilon_{jm})$ and $v_j(\mathbf{A}_j, \mathbf{D}_j) = \sum_{i=1}^m v_{ij}(\mathbf{A}_j, \mathbf{D}_j)$. Thus, for all the atoms $\underline{x}_j | \mathbf{A}_j$, we have that

$$P_j(\underline{x}_j | \mathbf{A}_j, g_{1j}, \dots, g_{mj}) = p_j(\underline{x}_j | \mathbf{A}_j) \prod_{i=1}^m g_{ij}^{v_{ij}(\mathbf{A}_j)}$$

for all $0 < g_{1j}, \dots, g_{mj} < 1/\mu_j^*(\underline{x}_j | \mathbf{A}_j)$. This implies that, for all j , (11) holds. \square

We are now in position to prove Theorem 4.2.

A.2. The proof of Theorem 4.2. If (Ω_j, μ_j^*) contains no atoms at all then the proof is immediate from Theorem A.1.

If (Ω_j, μ_j^*) is purely atomic, then (8) is easily obtained from (11) with $w_{ij}(\mathbf{A}_j) = v_{ij}(\mathbf{A}_j)$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$. Moreover, from the fact that f_j is CEB, it is easy to verify that the weights must sum to 1 for each j , that is, $\sum_{i=1}^m w_{ij}(\mathbf{A}_j) = 1$.

If (Ω_j, μ_j^*) has atoms but is not purely atomic, we use Theorem A.1 to obtain the result on the set N , the complement of the atoms set of (Ω_j, μ_j^*) . Consider the atoms $\underline{x}_j^{(1)} | \mathbf{A}_j^{(1)}, \underline{x}_j^{(2)} | \mathbf{A}_j^{(2)}, \dots$ in Ω_j and let $P_{kj}(g_{1j}, \dots, g_{mj})$ denote $P_j(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}, g_{1j}, \dots, g_{mj})$ for all $g_{ij}, i = 1, \dots, m, j = 1, \dots, n$, such that $0 < g_{ij} < 1/\mu_j^*(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)})$. From (7) in the definition of CEB for a PCG \mathcal{E}^{pd} we have for each j that

$$(12) \quad \frac{l_j(\underline{x}_j | \mathbf{A}_j)P_j[\underline{x}_j | \mathbf{A}_j, f_{1j}(\underline{x}_j | \mathbf{A}_j), \dots, f_{mj}(\underline{x}_j | \mathbf{A}_j)]}{P_j[\underline{x}_j | \mathbf{A}_j, h_{1j}(\underline{x}_j | \mathbf{A}_j), \dots, h_{mj}(\underline{x}_j | \mathbf{A}_j)]} = \text{constant},$$

μ_j^* -a.e.,

whenever $h_{ij} \propto l_j f_{ij}$ for all i and j .

From (10), we have that for $\underline{x}_j | \mathbf{A}_j$ on N the left-hand side of (12) equals $\prod_{i=1}^m t_{ij}^{w_{ij}(\mathbf{A}_j)}$, where $t_{ij} = \int \dots \int l_j f_{ij} d\mu_{1j} \dots d\mu_{r(j)j}$ for all i and j .

Fix t_{1j}, \dots, t_{mj} , pick a single vector of atoms $\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}$ in Ω_j and let ε_j be small enough such that $0 < \varepsilon_j/t_{ij} < 1/\mu_j^*(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)})$ for each i and j . Then, set $s_{ij} = f_{ij}(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}) = \varepsilon_j$ and $s_{ij}^* = h_{ij}(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}) = \varepsilon_j/t_{ij}$ for each i and j . For another vector $\underline{x}_j^{(q)} | \mathbf{A}_j^{(q)}$, set $u_{ij} = f_{ij}(\underline{x}_j^{(q)} | \mathbf{A}_j^{(q)})$ and $u_{ij}^* = h_{ij}(\underline{x}_j^{(q)} | \mathbf{A}_j^{(q)})$. Now, choose four sets in Ω_j for each j , say, B_{vj} , with $0 < \mu_j^*(B_{vj}) < \infty$ for $v = 1, 2, 3$ and $\mu_j^*(B_{4j}) > 0$, such that $s_{ij} \mu_j^*(B_{1j}) + u_{ij} \mu_j^*(B_{2j}) < 1$ and $s_{ij}^* \mu_j^*(B_{1j}) + u_{ij}^* \mu_j^*(B_{2j}) < 1$ for all i and j . Make $B_{1j} = \{\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}\}$ and $B_{2j} = \{\underline{x}_j^{(q)} | \mathbf{A}_j^{(q)}\}$ and construct densities $(f'_{1j}, \dots, f'_{mj})$ and likelihoods l_j for these densities such that $f'_{ij}(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}) = s_{ij}$, $f'_{ij}(\underline{x}_j^{(q)} | \mathbf{A}_j^{(q)}) = u_{ij}$, $h'_{ij}(\underline{x}_j^{(k)} | \mathbf{A}_j^{(k)}) = s_{ij}^*$ and $h'_{ij}(\underline{x}_j^{(q)} | \mathbf{A}_j^{(q)}) = u_{ij}^*$, where $h'_{ij} = l_j f'_{ij} / \int \dots \int l_j f'_{ij} d\mu_{1j} \dots d\mu_{mj}$ (for all i and j).

Denote $\gamma_{ij} = s_{ij}\mu_j^*(B_{1j}) + u_{ij}\mu_j^*(B_{2j})$. Observe that $s_{ij}\mu_j^*(B_{1j}) + u_{ij}\mu_j^*(B_{2j}) = t_{ij}[s_{ij}^*\mu_j^*(B_{1j}) + u_{ij}^*\mu_j^*(B_{2j})] < t_{ij}$ for all i and j , and thus $0 < \gamma_{ij} < \min\{t_{ij}, 1\}$ for each i and j . Also choose $\lambda_j > 0$ and $\xi_j < \infty$ such that

$$\begin{aligned} \lambda_j &< \min\left\{[1 - \gamma_{ij}]^{-1}(t_{ij} - \gamma_{ij}): i = 1, \dots, n\right\} \\ &\geq \max\left\{[1 - \gamma_{ij}]^{-1}(t_{ij} - \gamma_{ij}): i = 1, \dots, n\right\} < \xi_j. \end{aligned}$$

Fixing an arbitrary density $z_j \in \Delta$ we can define, for all i and j ,

$$\begin{aligned} f'_{ij} &= s_{ij}\mathcal{S}(B_{1j}) + u_{ij}\mathcal{S}(B_{2j}) + \frac{t_{ij} - \gamma_{ij} - \lambda_j(1 - \gamma_{ij})}{\mu_j^*(B_{3j})(\xi_j - \lambda_j)}\mathcal{S}(B_{4j}) \\ &\quad + \frac{\xi_j(1 - \gamma_{ij}) - (t_{ij} - \gamma_{ij})}{R_j(\xi_j - \lambda_j)}z_j\mathcal{S}(B_{4j}), \end{aligned}$$

where $R_j = \int \cdots \int \mathcal{S}(B_{4j})z_j d\mu_{1j} \cdots d\mu_{r(j)j} > 0$ and $\mathcal{S}(B)$ is the indicator of the set B . Note that $\int \cdots \int f'_{ij} d\mu_{1j} \cdots d\mu_{r(j)j} = 1$ and $f'_{ij}(x_j^{(k)} | \mathbf{A}_j^{(k)}) = s_{ij}$, $f'_{ij}(x_j^{(q)} | \mathbf{A}_j^{(q)}) = u_{ij}$ for all i and j . Finally, consider the likelihoods

$$l_j = \mathcal{S}(B_{1j}) + \mathcal{S}(B_{2j}) + \xi_j\mathcal{S}(B_{3j}) + \lambda_j\mathcal{S}(B_{4j}).$$

Using the fact that (12) also holds on all of Ω_j , we have

$$\frac{P_{kj}(\varepsilon_j, \dots, \varepsilon_j)}{P_{kj}(\varepsilon_j/t_{1j}, \dots, \varepsilon_j/t_{mj})} = \prod_{i=1}^m t_{ij}^{w_{ij}(\mathbf{A}_j^{(k)})}$$

for each j and an arbitrary k . This means that, for all g_{1j}, \dots, g_{mj} strictly between 0 and $1/\mu_j^*(x_j^{(k)} | \mathbf{A}_j^{(k)})$,

$$P_{kj}(g_{1j}, \dots, g_{mj}) = P_{kj}(\varepsilon_j, \dots, \varepsilon_j) \varepsilon_j^{-1} \prod_{i=1}^m g_{ij}^{w_{ij}(\mathbf{A}_j^{(k)})}.$$

Let $P_j(x_j^{(k)} | \mathbf{A}_j^{(k)}) = P_{kj}(\varepsilon_j, \dots, \varepsilon_j) \varepsilon_j^{-1}$ and (8) is proved.

Although the weights can possibly vary with \mathbf{A}_j , they are constant relative to $x_j | \mathbf{A}_j$ for fixed values of \mathbf{A}_j , $j = 1, \dots, n$. Also, they must be nonnegative if μ_j^* is not purely atomic or if it is purely atomic but Ω_j includes a countably infinite number of atoms. In this latter case, we can easily construct densities f_{1j}, \dots, f_{mj} which will make the integrals

$$\int \cdots \int p_j \prod_{i=1}^m f_{ij}^{w_{ij}(\mathbf{A}_j)} d\mu_{1j} \cdots d\mu_{r(j)j}$$

infinite unless all the weights are nonnegative. However, these integrals are always finite when Ω_j is finite and μ_j^* is a counting-type measure. The weights can take negative values in this case. Furthermore, p_j must be essentially bounded, or else there exist f_j such that the above integrals are infinite when all the f_{ij} are equal to f_j for each j , according to Hewitt and Stromberg [(1965), Theorem 20.15].

APPENDIX B

PROOF OF THEOREM 5.1. Here we use the convention that a function will explicitly depend only on the values of its arguments. For simplicity, the argument of w_{ij} , that is $\pi(\underline{x}_k)$, will be omitted in this proof.

First, observe that there is a proportionality constant for each $k = 1, \dots, n$, $h_k[\pi(\underline{x}_k), \underline{z}^k] = 1/f \cdots \int \prod_{i=1}^m f_{ij}^{w_{ik}}[\underline{x}_k | \pi(\underline{x}_k), \underline{z}^k] dx_{1k} \cdots dx_{r(k)k}$, $r(k)$ being the dimension of the vector element \underline{X}_k , such that

$$\tilde{f}_k[\underline{x}_k | \pi(\underline{x}_k), \underline{z}^k] = h_k[\pi(\underline{x}_k), \underline{z}] \prod_{i=1}^m f_{ik}^{w_{ik}}[\underline{x}_k | \pi(\underline{x}_k), \underline{z}^k],$$

where $\underline{Z}^k = (\underline{Z}_1, \dots, \underline{Z}_k)$. Also, if information \underline{Z}^k about \underline{X}^k is observed then the density of $(\underline{X}_{k+1}, \dots, \underline{X}_n)$ given \underline{X}^k remains unchanged.

We begin the proof by showing that, under the conditions of the theorem, the density of $\underline{X}_n | \{\pi(\underline{X}_n), \underline{Z}^n\}$ does not depend on when \underline{Z}^n is incorporated. Notice that $\underline{Z}^n = \underline{Z}$.

Since $g(\underline{z} | \underline{x})$ is a function of \underline{x}_n only through $g_n[\underline{z}_n | \underline{x}_n, \pi(\underline{x}_n), \underline{z}^{n-1}]$, we can write that

$$\begin{aligned} f_n[\underline{x} | \pi(\underline{x}_n), \underline{z}^n] &= h_n[\pi(\underline{x}_n), \underline{z}^n] \\ &\times \prod_{i=1}^m \left[\frac{f_{in}[\underline{x}_n | \pi(\underline{x}_n)] g_n[\underline{z}_n | \pi(\underline{x}_n), \underline{x}_n, \underline{z}^{n-1}]}{h_{in}[\pi(\underline{x}_n), \underline{z}^n]} \right]^{w_{in}}, \end{aligned}$$

where throughout we let

$$\begin{aligned} h_{in}[\pi(\underline{x}_n), \underline{z}^n] &= \int \cdots \int f_{in}[\underline{x}_n | \pi(\underline{x}_n)] \\ &\times g_n[\underline{z}_n | \underline{x}_n, \pi(\underline{x}_n), \underline{z}^{n-1}] dx_{1n} \cdots dx_{r(n)n} \end{aligned}$$

for $i = 1, \dots, m$. Noting that $\sum_{i=1}^m w_{in} = 1$, this can be arranged as

$$\begin{aligned} \tilde{f}_n[\underline{x}_n | \pi(\underline{x}_n), \underline{z}^n] &= \frac{h_n[\pi(\underline{x}_n), \underline{z}^n]}{\prod_{i=1}^m h_{in}^{w_{in}}[\pi(\underline{x}_n), \underline{Z}^n]} \\ &\times \prod_{i=1}^m f_{in}^{w_{in}}[\underline{x}_n | \pi(\underline{x}_n)] g_n[\underline{z}_n | \underline{x}_n, \pi(\underline{x}_n), \underline{z}^{n-1}] \\ &= V_n[\pi(\underline{x}_n), \underline{z}^n] f_n[\underline{x}_n | \pi(\underline{x}_n), \underline{z}^n], \end{aligned}$$

where

$$V_n[\pi(\underline{x}_n), \underline{z}^n] = \frac{h_n[\pi(\underline{x}_n), \underline{z}^n] v_n[\pi(\underline{x}_n), \underline{z}^n]}{\prod_{i=1}^m h_{in}^{w_{in}}[\pi(\underline{x}_n), \underline{z}^n]}$$

and where

$$\begin{aligned} v_n[\pi(\underline{x}_n), \underline{z}^n] &= \\ &= \frac{1}{\int \cdots \int \prod_{i=1}^m f_{in}^{w_{in}}[\underline{x}_n | \pi(\underline{x}_n)] g_n[\underline{z}_n | \underline{x}_n, \pi(\underline{x}_n), \underline{z}^{n-1}] dx_{1n} \cdots dx_{r(n)n}}. \end{aligned}$$

Now we know that \tilde{f}_n and \bar{f}_n must integrate to 1 over \underline{x}_n for all values of $\pi(\underline{x}_n)$ and \underline{z}^n . It follows therefore that $V_n[\pi(\underline{x}_n), \underline{z}^n]$ is identically 1 and so

$$(13) \quad \tilde{f}_n[\underline{x}_n | \pi(\underline{x}_n), \underline{z}^n] = \bar{f}_n[\underline{x}_n | \pi(\underline{x}_n), \underline{z}^n]$$

as required for each node of the ID induced by \mathcal{G}^{pd} .

After having shown that whether the combination is done before or after observing \underline{Z} does not affect the conditional density of $\underline{X}_n | \{\pi(\underline{X}_n), \underline{Z}^n\}$, we next consider the updating of $\underline{X}_{n-1} | \pi(\underline{X}_{n-1})$ given \underline{Z}^n . First note that to update the distribution of \underline{X}^{n-1} , and hence of $\underline{X}_{n-1} | \pi(\underline{X}_{n-1})$, in the light of \underline{Z}^n , we need to calculate the density $g^{(1)}$ of \underline{z}^n given \underline{x}^{n-1} . We can then simply use Bayes' rule as above and since the likelihood of \underline{Z} is cutting, from the usual probability calculus we can write that

$$g^{(1)}[\underline{z}^n | \underline{x}^{n-1}] = g_1(z_1 | \underline{x}_1) g_2[z_2 | \underline{x}_2, \pi(\underline{x}_2), z_1] \\ \times \cdots g_{n-1}[z_{n-1} | \underline{x}_{n-1}, \pi(\underline{x}_{n-1}), \underline{z}^{n-2}] \tilde{g}_n(z_n | \underline{x}^{n-1}, \underline{z}^{n-1}),$$

where

$$\tilde{g}_n(z_n | \underline{x}^{n-1}, \underline{z}^{n-1}) = \int \cdots \int g_n[z_n | \underline{x}_n, \pi(\underline{x}_n), \underline{z}^{n-1}] \\ \times f_n[\underline{x}_n | \pi(\underline{x}_n)] dx_{1n} \cdots dx_{r(n)n}.$$

Note that f_n above is unambiguously defined since $\tilde{f}_n = \bar{f}_n$ by (13) and

$$f_n[\underline{x}_n | \pi(\underline{x}_n)] = \bar{h}[\pi(\underline{x}_n)] \prod_{i=1}^m f_{in}^{w_{in}}[\underline{x}_n | \pi(\underline{x}_n)].$$

Now, provided that w_{in} is a function of \underline{x} only of terms in $\pi(\underline{x}_n)$, it is clear that \tilde{g}_n is a function of \underline{z}^n and $\pi(\underline{x}_n)$ only. So we can write that

$$\tilde{g}_n(z_n | \underline{x}^{n-1}, \underline{z}^{n-1}) = \tilde{g}_n[z_n | \pi(\underline{x}_n), \underline{z}^{n-1}],$$

where \tilde{g}_n is a function of its arguments only. Also, since $\mathcal{G}^{pd}(\underline{X})$ has a decomposable induced ID, it will exhibit the *running intersection property* [Lauritzen and Spiegelhalter (1988)]. This states that there will exist an index $j(n)$ (say) such that $\pi(\underline{x}_n) \subseteq \{\underline{x}_{j(n)}, \pi(\underline{x}_{j(n)})\}$ with $j(n) = 1, \dots, n - 1$. So \tilde{g} can be further simplified into

$$\tilde{g}_n(z_n | \underline{x}^{n-1}, \underline{z}^{n-1}) = \tilde{g}_n[z_n | \underline{x}_{j(n)}, \pi(\underline{x}_{j(n)}), \underline{z}^{n-1}].$$

Hence

$$g^{(1)}[\underline{z}^n | \underline{x}^{n-1}] = g_1^{(1)}[z_1^{(1)} | \underline{x}_1] g_2^{(1)}[z_2^{(1)} | \underline{x}_2, z_1^{(1)}] \\ \times \cdots g_{n-1}^{(1)}[z_{n-1}^{(1)} | \underline{x}_{n-1}, \pi(\underline{x}_{n-1}), \underline{z}^{(1), n-2}],$$

where

$$g_j^{(1)}[z_j^{(1)} | \underline{x}_j, \pi(\underline{x}_j), \underline{z}^{(1), j-1}] = g_j[z_j^{(1)} | \underline{x}_j, \pi(\underline{x}_j), \underline{z}^{(1), j-1}],$$

and $z_k^{(1)} = z_k$ for $k \neq j(n)$, $k = 1, \dots, n - 1$ and

$$g_{j(n)}^{(1)}[z_{j(n)}^{(1)} | \underline{x}_{j(n)}, \pi(\underline{x}_{j(n)}), \underline{z}^{j(n)-1}] \\ = g_{j(n)}[z_{j(n)} | \underline{x}_{j(n)}, \pi(\underline{x}_{j(n)}), \underline{z}^{j(n)-1}] \tilde{g}_n[z_n | \underline{x}_{j(n)}, \pi(\underline{x}_{j(n)}), \underline{z}^{n-1}].$$

It follows that $g^{(1)}(\underline{z} | \underline{x}^{n-1})$ is a likelihood function which is cutting on $\underline{x}^{(1)} = \underline{x}^{n-1}$ since the only changed term is the likelihood arising as if from two conditionally independent observations $\underline{z}_j^{(1)} = (\underline{z}_{j(n)}, \underline{z}_n)$.

Now consider the density of $\underline{X}_{n-1} | \{\pi(\underline{X}_{n-1}), \underline{Z}\}$. The argument leading to (13) can be applied with $n - 1$ replacing n , g replaced by $g^{(1)}$ and \underline{x} by $\underline{x}^{(1)}$ to give us that

$$\tilde{f}_{n-1}[\underline{x}_{n-1} | \pi(\underline{x}_{n-1}), \underline{z}^{n-1}] = \tilde{f}[\underline{x}_{n-1} | \pi(\underline{x}_{n-1}), \underline{z}^{n-1}].$$

Since the induced ID is decomposable, we can find an index $j(n - 1)$ such that $\pi(\underline{x}_{n-1}) \subseteq \{\underline{x}_{j(n-1)}, \pi(\underline{x}_{j(n-1)})\}$. We can therefore use an argument exactly analogous to the one above, replacing $\underline{x}^{(1)}$ by $\underline{x}^{(2)} = \underline{x}^{n-2} = (\underline{x}_1, \dots, \underline{x}_{n-2})$, $(n - 1)$ by $(n - 2)$, g by $g^{(1)}$ and $g^{(1)}$ by $g^{(2)}$, where

$$g^{(2)}[\underline{z}^{n-1} | \underline{x}^{n-2}] = g_1^{(2)}[\underline{z}_1^{(2)} | \underline{x}_1] \cdots g_{n-2}^{(2)}[\underline{z}_{n-2}^{(2)} | \underline{x}_{n-2}, \pi(\underline{x}_{n-2}), \underline{z}^{(2), n-3}]$$

with

$$g_j^{(2)}[\underline{z}_j^{(2)} | \underline{x}_j, \pi(\underline{x}_j), \underline{z}^{(2), j-1}] = g_j[\underline{z}_j^{(2)} | \underline{x}_j, \pi(\underline{x}_j), \underline{z}^{(2), j-1}]$$

and $\underline{z}_j^{(2)} = \underline{z}_j^{(1)}$ for $j \neq j(n - 1)$, $j = 1, \dots, n - 2$. Also,

$$\begin{aligned} g_{j(n-1)}^{(2)}[\underline{z}_{j(n-1)}^{(2)} | \underline{x}_{j(n-1)}, \pi(\underline{x}_{j(n-1)}), \underline{z}^{(2), j(n-1)-1}] \\ = g_{j(n-1)}^{(1)}[\underline{z}_{j(n-1)}^{(1)} | \underline{x}_{j(n-1)}, \pi(\underline{x}_{j(n-1)}), \underline{z}^{j(n-1)-1}] \\ \times g_{n-1}^{(1)}[\underline{z}_{n-1}^{(1)} | \pi(\underline{x}_{j(n-1)}), \underline{z}^{(2), j(n-1)}] \end{aligned}$$

and we have that

$$\tilde{f}_{n-2}[\underline{x}_{n-2} | \pi(\underline{x}_{n-2}), \underline{z}^{n-2}] = \tilde{f}_{n-2}[\underline{x}_{n-2} | \pi(\underline{x}_{n-2}), \underline{z}^{n-1}].$$

It should now be clear that we can proceed inductively backwards through the indices k , starting from $k = n$ to prove that $\tilde{f}_k[\underline{x}_k | \pi(\underline{x}_k), \underline{z}^k] = \tilde{f}_k[\underline{x}_k | \pi(\underline{x}_k), \underline{z}^k]$ for $k = 1, \dots, n$. However, since the density of \tilde{f} is uniquely determined by the product of \tilde{f}_k and f by the product of conditional densities \tilde{f}_k , $k = 1, \dots, n$, we have then proved that

$$\tilde{f}(x | z) = \tilde{f}(\underline{x} | \underline{z}),$$

that is, a *sequential* external Bayesianity holds for $\mathcal{E}^{pd}(\underline{X} | \underline{Z})$. \square

Acknowledgments. We are grateful to two referees and an Associate Editor for their helpful comments and suggestions on the first version of the paper.

REFERENCES

BACHARACH, M. (1972). Scientific disagreement. Christchurch, Oxford. Unpublished manuscript.
 BACHARACH, M. (1975). Group decisions in the face of differences of opinion. *Management Science* **22** 182–191.
 BATES, J. M. and GRANGER, C. W. J. (1969). The combination of forecasts. *Oper. Res.* **20** 451–468.

- BAYARRI, M. J. and DEGROOT, M. H. (1988). Gaining weight: a Bayesian approach. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 25–44. Oxford Univ. Press.
- COOKE, R. M. (1991). *Experts in Uncertainty: Expert Opinion and Subjective Probability in Science*. Oxford Univ. Press.
- DALKEY, N. C. (1972). An impossibility theorem for group probability functions. Report No. P-4862, The Rand Corporation, Santa Monica, CA.
- DALKEY, N. C. (1975). Toward a theory of group estimation. In *The Delphi Method: Techniques and Applications* (H. A. Linstone and M. Turoff, eds.) 236–261. Addison-Wesley, Reading, MA.
- DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 1–31.
- DAWID, A. P. (1992). Applications of a general propagation algorithm for probabilistic expert systems. *Statist. Comput.* **2** 25–36.
- DEGROOT, M. H. and MORTERA, J. (1991). Optimal linear opinion pools. *Management Science* **37** 546–558.
- FARIA, A. E. (1996). Graphical Bayesian models in multivariate expert judgements and conditional external Bayesianity. Ph.D. dissertation, Dept. Statistics, Univ. Warwick.
- FARIA, A. E. and SMITH, J. Q. (1994). Conditional external Bayesianity in chain graphs. Technical Report 271, Dept. Statistics, Univ. Warwick.
- FARIA, A. E. and SMITH, J. Q. (1996). Conditional external Bayesianity in decomposable influence diagrams. In *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 551–560. Clarendon Press, Oxford.
- FARIA, A. E. and SOUZA, R. C. (1995). A re-evaluation of the quasi-Bayes approach to the linear combination of forecasts. *Journal of Forecasting* **14** 533–542.
- FRENCH, S. (1985). Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 183–202. North-Holland, Amsterdam.
- FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Statist.* **17** 333–353.
- GEANOKOPIOS, J. (1992). Common knowledge. *Journal of Economic Perspectives* **6** 53–82.
- GENEST, C. (1984). A characterization theorem for externally Bayesian groups. *Ann. Statist.* **12** 1110–1115.
- GENEST, C., MCCONWAY, K. J. and SCHERVISH, M. J. (1986). Characterization of externally Bayesian pooling operators. *Ann. Statist.* **14** 487–501.
- GENEST, C. and WAGNER, C. G. (1984). Further evidence against independence preservation in expert judgement synthesis. Technical Report 84-10, Dept. Statistics and Actuarial Science, Univ. Waterloo.
- GENEST, C. and ZIDEK, J. V. (1986). Combining probability distributions: a critique and an annotated bibliography (with discussion). *Statist. Sci.* **1** 114–148.
- HEWITT, E. and STROMBERG, K. (1965). *Real and Abstract Analysis*. Springer, New York.
- JENSEN, F., JENSEN, F. V. and DITTMER, S. L. (1994). From influence diagrams to junction trees. Report R-94-2013, Dept. Mathematics and Computer Science, Aalborg Univ., Denmark.
- LADDAGA, R. (1977). Lehrer and the consensus proposal. *Synthese* **36** 473–477.
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Univ. Press.
- LAURITZEN, S. L., SPEED, T. P. and VIJAYAN, K. (1984). Decomposable graphs and hypergraphs. *J. Austral. Math. Soc. Ser. A* **36** 12–29.
- LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50** 157–224.
- LEHRER, K. and WAGNER, C. G. (1983). Probability amalgamation and the independence issue: a reply to Laddaga. *Synthese* **55** 339–346.
- LINDLEY, D. V. (1985). Reconciliation of discrete probability distributions. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 375–390. North-Holland, Amsterdam.
- MADANSKY, A. (1964). Externally Bayesian groups. Rand Memo RM-4141-PR.

- MADANSKY, A. (1978). Externally Bayesian groups. Univ. Chicago. Unpublished manuscript.
- MCCONWAY, K. J. (1978). The combination of experts' opinions in probability assessment: some theoretical considerations. Ph.D. dissertation, Univ. College London.
- MCCONWAY, K. J. (1981). Marginalization and linear opinion pools. *J. Amer. Statist. Assoc.* **76** 410–414.
- MENDEL, M. B. and SHERIDAN, T. B. (1987). Optimal estimation using human experts. Man-machine Systems Lab., Dept. Mechanical Engineering, MIT.
- PEARL, J. (1988). *Probabilistic Reasoning in Intelligence Systems*. Morgan Kaufmann, San Mateo, CA.
- PILL, J. (1971). The Delphi method: substance, context, a critique and an annotated bibliography. *Socio-Economic Planning and Science* **5** 57–71.
- QUEEN, C. M. and SMITH, J. Q. (1993). Multiregression dynamic models. *J. Roy. Statist. Soc. Ser. B* **55** 849–870.
- RAIFFA, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty* 211–226. Addison-Wesley, Reading, MA.
- RUDIN, W. (1986). *Real and Complex Analysis*. McGraw-Hill, New York.
- SMITH, A. F. M. and MAKOV, U. E. (1978). A quasi-Bayes sequential procedure for mixtures. *J. Roy. Statist. Soc. Ser. B* **40** 106–112.
- SMITH, J. Q. (1989). Influence diagrams for statistical modelling. *Ann. Statist.* **17** 654–672.
- SMITH, J. Q. and PAPAMICHAIL, K. N. (1996). Fast Bayes and the dynamic junction forest. Real Time Online Decision Support System document (WG1)-TN(96)-02.
- STONE, M. (1961). The opinion pool. *Ann. Math. Statist.* **32** 1339–1342.
- WERMUTH, N. and LAURITZEN, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. Roy. Statist. Soc. Ser. B* **52** 21–50.
- WEST, M. and HARRISON, P. J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer, New York.
- WINKLER, R. L. (1968). The consensus of subjective probability distributions. *Management Science* **15** B61–B75.
- WINKLER, R. L. (1986). Expert resolution. *Management Science* **32** 298–303.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
ENGLAND