Coherent Inference on Distributed Bayesian Expert Systems

Jim Q. Smith

Warwick University

Sep 2011

Abstract

It is becoming increasingly necessary for different probabilistic expert systems to be networked together. Different collections of domain experts must independently specify their judgments within each component system and update these in the light of the data they receive. But in these circumstances what overarching beliefs must the collective agree and what types of data can be admitted in the system so that the collective acts as if it were a single Bayesian? In this talk I will explore these issues and illustrate the main technical problems through discussing some simple examples.

The Setting

Decision Support for a single Bayesian user:

- User adopts expert judgments as her own.

The Setting

Decision Support for a single Bayesian user:

- User adopts expert judgments as her own.
- Network of different panels of experts over different domains.

The Setting

Decision Support for a single Bayesian user:

- User adopts expert judgments as her own.
- Network of different panels of experts over different domains.
- On-line updating necessary.

The Setting

Decision Support for a single Bayesian user:

- User adopts expert judgments as her own.
- Network of different panels of experts over different domains.
- On-line updating necessary.
- Coherence and auditability.

So more specifically

The Decision Support system has:

- Large number of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.

So more specifically

The Decision Support system has:

- Large number of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.
- Different, panels $\{G_1, G_2, \ldots G_m\}$ of domain experts (the collective) oversee different domains.

So more specifically

The Decision Support system has:

- Large number of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.
- Different, panels $\{G_1, G_2, \ldots G_m\}$ of domain experts (the collective) oversee different domains.
- Agreed qualitative framework used to paste judgments into a single probability model.

So more specifically

The Decision Support system has:

- Large number of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.
- Different, panels $\{G_1, G_2, \ldots G_m\}$ of domain experts (the collective) oversee different domains.
- Agreed qualitative framework used to paste judgments into a single probability model.
- User's prespecified class of utility functions can help simplify required inputs.

So more specifically

The Decision Support system has:

- Large number of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.
- Different, panels $\{G_1, G_2, \ldots G_m\}$ of domain experts (the collective) oversee different domains.
- Agreed qualitative framework used to paste judgments into a single probability model.
- User's prespecified class of utility functions can help simplify required inputs.
- Support: identifies and explains user's expected utility maximising decisions.

So more specifically

The Decision Support system has:

- Large number of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$.
- Different, panels $\{G_1, G_2, \ldots G_m\}$ of domain experts (the collective) oversee different domains.
- Agreed qualitative framework used to paste judgments into a single probability model.
- User's prespecified class of utility functions can help simplify required inputs.
- Support: identifies and explains user's expected utility maximising decisions.
- All adaptations to admissible data must appear rational from the outside.

Example: decision support after a nuclear accident

Many panels of experts/statistical models in the system:

- Power station described by a Bayesian Network - **Panel** nuclear physicists, engineers and managers.
- Accidental release into the atmosphere or water supply the dangerous radiation will be distributed into the environment, **Panel** atmospheric physicists, hydrologist, local weather forecasters....
- Taking outputs of dispersion models and data on demography and implemented countermeasures predict exposure of humans animal and plants of the contaminant. **Panel** biologists Food scientists, local adminstrators, ..
- Taking outputs giving type and extent of exposure predict health consequences: **Panel** epidemiologists, medics, genetic researchers
- And so on ...

So more formally

- *Collective* jointly responsible for all the probability statements for intirinsic vector $\mathbf{Y}$.informing potential user's *reward* vector $\mathbf{R}$ - of her utility. ($\mathbf{Y}(\mathbf{R})$ often indexed by $d \in D$)
- Each panel G_i, $i = 1.2, \ldots, m$ delivers beliefs $\{\Pi_i(d) : d \in D\}$.about the parameters of $P(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{z}_i, d)$, where $\mathbf{Y}_i(d), \mathbf{Z}_i(d)$ are disjoint ($\mathbf{Z}_i(d)$ possibly null) subvectors of $\mathbf{Y}(d)$.
- Call Θ_i the *domain*, $\Pi_i(d)$ the *panel beliefs* ($\pi_i(\boldsymbol{\theta}_i, d)$ the *panel density*)

Key point: each panel *only* provides collective with quantative (composite) beliefs concerning *their particular* domain.

Example: Observables a pair of binary variables

- $\mathbf{R} = \mathbf{Y} \triangleq (Y_1, Y_2)$. Panel G_1 inputs about $\theta_1 \triangleq P(Y_1 = 1)$.
- Panel G_2, $\theta_{2,0} \triangleq P(Y_2 = 1 | Y_1 = 0)$ and $\theta_{2,1} \triangleq P(Y_2 = 0 | Y_1 = 1)$.
- Distribution of $\mathbf{R}$, $\overline{\boldsymbol{\theta}} \triangleq \left(\overline{\theta}_{00}, \overline{\theta}_{01}, \overline{\theta}_{10}, \overline{\theta}_{11}\right)$ given by the polynomials

$$\begin{aligned} \overline{\theta}_{00} &= (1-\theta_1)(1-\theta_{2,0}), \overline{\theta}_{01} = (1-\theta_1)\theta_{2,0}, \\ \overline{\theta}_{10} &= \theta_1(1-\theta_{2,1}), \overline{\theta}_{11} = \theta_1\theta_{2,1} \end{aligned}$$

- G_1 donates densities $\Pi_1 = \{\pi_1(\theta_1, d) : d \in D\}$.
- G_2 gives densities $\Pi_2 = \{(\pi_2(\theta_{2,0}, d), \pi_2(\theta_{2,1}, d)) : d \in D\}$.

Recapping the Problem

- Collective agrees set of qualitative (e.g. conditional independence) assumptions about $\{Y_i : 1 \leq i \leq n\}$ conditional on $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_m)$ whatever $d \in D$.
- Let $\Pi = f(\Pi_1, \Pi_2, \ldots, \Pi_m)$ be the distributional statements about $\boldsymbol{\theta}$ available to the user. Panel beliefs $\{\Pi_j(d) : 1 \leq j \leq m, d \in D\}$ the *only* quantitative inputs to the collective beliefs $\Pi(d)$ about $\boldsymbol{\theta}$.

Note: not trivial that $\Pi(d)$ is function of $\Pi_j(d) : 1 \leq j \leq m$.
e.g distribution of parameters of $\mathbf{Y} = (Y_1, Y_2)$ is not fully recoverable from the two marginal densities $\pi_i(\boldsymbol{\theta}_i)$,provided by G_i, $i = 1, 2$ e.g. no covariance between Y_1 and Y_2 .

Questions to Answer

1. When and how can panel judgments be combined to provide a *coherent composite system*?
2. Given Π is sufficiently detailed and coherent what *protocols* need to be followed? When does $\pi(\overline{\boldsymbol{\theta}})$ define the *genuine* beliefs held by the collective and user?
3. For online distributed updating, panels must *update their beliefs autonomously* with the data available to provide individual inputs $\{\Pi_i. : 1 \leq i \leq m\}$.to a new coherent specification within the same framework. What beliefs must the collective share about accommodated data structures for f to respect this updating? What characteristics of *admissible data* makes this possible?

We will see that such a system is surprisingly easy to define if we *restrict* data allowed.

Example: The Queen in Danger!!

Example

Panel G_1 domain is margin of binary Y_1 - $\theta_1 = P(Y_1 = 1)$ (Y_1 queen comes in contact with a particular virus). Panel G_2 domain margin of binary Y_2, $\theta_2 = P(Y_2 = 1)$. (Y_2 when queen exposed suffers an adverse reaction).G_1 says $\theta_1 \backsim Be(\alpha_1, \beta_1)$ and G_2 says $\theta_2 \backsim Be(\alpha_2, \beta_2)$. No decision will affect these distributions. Agreed structural information is $Y_1 \amalg Y_2 | (\theta_1, \theta_2)$,

Case1: User has a separable utility

$$u_1(y_1, y_2, d_1, d_2) = a + b_1(d_1)y_1 + b_2(d_2)y_2$$

G_i needs only supply $\mu_i \triangleq \mathbb{E}(\theta_i) = \alpha_i(\alpha_i + \beta_i)^{-1}$, $i = 1, 2$. No need to be concerned about dependency.

Example

Case 2

- Interest is only in $W \triangleq Y_1 Y_2$ (whether queen is infected). So

$$u_2(w, d_{12}) = a + b_{12}(d_{12})w$$

where $\mathbb{E}(W) = \mathbb{E}(\theta_1\theta_2)$.

- If collective assumes *global independence* $\Rightarrow$ distribution $\theta_1\theta_2$ is well defined.
- Then $\mathbb{E}(\theta_1\theta_2) = \mu_1\mu_2$ - so G_i needs only supply μ_i, $i = 1, 2$.
- However Global independence not *only* choice!

An Alternative Prior

Suppose $\alpha_1 + \beta_1 = \alpha_2 + \beta_2 \triangleq \sigma$. Panels donate (μ_1, μ_2, σ), where $\sigma = \gamma_{00} + \gamma_{10} + \gamma_{10} + \gamma_{11}$, $\pi \backsim Di(\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11})$,

$$\begin{aligned} \alpha_1 &= \gamma_{10} + \gamma_{11}, \beta_1 = \gamma_{00} + \gamma_{01} \\ \alpha_2 &= \gamma_{01} + \gamma_{11}, \beta_2 = \gamma_{00} + \gamma_{10} \end{aligned}$$

- This collective prior consistent with panel margins but *not* global independence.
- Collective parameters $(\mu_1, \mu_2, \sigma, \rho)$, $\rho \triangleq \sigma^{-2} (\gamma_{11}\gamma_{00} - \gamma_{10}\gamma_{011})$
- Collective's $\mathbb{E}(\theta_1\theta_2) = \gamma_{11}\sigma^{-1} = \mu_1\mu_2 + \rho \neq \mu_1\mu_2$ unless $\rho = 0$.
- So $\mathbb{E}(\theta_1\theta_2)$ is not identified from inputs.

Now assume global independence

- Panels supplement judgments by independently randomly sampling.
- Collective needs only two updated posterior means $\mu_i^* . i = 1, 2$.
- So all data of this form allows distributed inference.

Problem 1: Global independence critical for distributivity. Even in Case 1 when only individuals margins of θ_1, θ_2 needed if collective did not believe $\theta_1 \amalg \theta_2$ it would need to draw on what it learns about θ_2 - through G_2's experiments to modify distribution of θ_1.
Problem 2 :Even if global independence is justified, assuming experiments of two panels never mutually informative also critical.

Example of data set: table of counts below (Case 2)

$Y_1 \backslash Y_2$	0	1		
0	5	45	50	$n - x_1$
1	45	5	50	x_1
	50	50	100	
	$n - x_2$	x_2		

- Each panel updates using only their respective margin (with weak priors) $\Rightarrow \mu_i^* \simeq 0.5$, $i = 1, 2 \Rightarrow \mathbb{E}(\theta_1\theta_2)$ to be approximately 0.25.
- OTOH with whole info $\mathbb{E}(\theta_1\theta_2) \simeq 0.05$.i.e. five times smaller!

(Note structural independence assumption: $Y_2 \amalg Y_1 | (\theta_1, \theta_2)$ looks dubious)

Non-compatible sampling

Binomial sample 100 units like queen, *acquiring* disease, so prob $\phi \triangleq P(W = 1)$. See 5 infected.

- In either case collective easily incorporates this information directly: e.g. giving ϕ a beta prior and treating data as random sample. However, without further assumptions such data impossible for G_i to *individually* update $\pi_i(\theta_i)$.
- Ignore this information $\div$ uniform priors $\Rightarrow$ vastly overestimate the probability.
- So $\pi(\theta_1\theta_2)$ no longer decomposes into a G_1 density and a G_2 density: Sampling induces dependence.

So even in simplest scenarios, problems quite involved! Need to be sensitive to what information is received.

External Bayesianity

External Bayesianity (EB) if all *individually* update priors using experiment (common knowledge) - giving likelihood $l(\boldsymbol{\theta}|\mathbf{x})$ - this same as if all first combined beliefs into single panel density to accommodate their new information and then updated.
EB property characterises the *logarithmic pool*

$$\overline{\pi}(\boldsymbol{\theta}|\mathbf{w}) \propto \prod_{i=1}^{k} \pi_i^{w_i}(\boldsymbol{\theta})$$

where $\mathbf{w} = (w_1, \ldots, w_k)$ weights, reflecting credibility of different experts, sum to unity.
Collective appears Bayesian from outside irrespective of sampling and order of information. Consistent with the Strong Likelihood Principle. Preserves integrity of panel independence over time.

Beliefs and Facts: What goes into system?

- Shared *beliefs* collective agrees reflect best (generally acceptable) available judgments about the global domain. Examples ci / causal/ functional relationships hardwired into system.
- Accepted *facts* Published data from well conducted experiments and sample surveys/events.

BUT most analyses implicitly or explicitly exclude certain data
Typical selection criteria:

- *Compellingness* of the evidence (e.g.to user ÷ auditor/Cochraine).
- *Defensibility* of modeling assumptions needed to be employed.
- *Wealth* of less ambiguous and less costly evidence

Held v Stated Bayesian beliefs Collective updates *only* in the light of agreed experiments/surveys/observational studies . Cannot use *all* relevant information.

Comments about what to include in an analysis

- Any practical Bayesian expert system needs a protocol for what information is *admitted* into the system.
- Such an *admissibility protocol* decided before seeing data $\mathbf{x}_t$ from a collection of experiments (sample surveys observational studies) $\mathcal{E}_t$ will be available to the collective at time t,

Information not incorporated still useful e.g. for diagnostics.

- An admissiblity protocol has the *separability property* if it only admits data $\mathbf{x}_t$ to time t whose associated likelihood is panel separable.

Separable Likelihoods: The key to distributivity

Definition

A set of experiments $\mathcal{E}$ with likelihood $l(\boldsymbol{\theta}|\mathbf{x},d)$,$d \in D$, is *panel separable* over $\boldsymbol{\theta}_i$, $i = 1, \ldots, m$ when

$$l(\boldsymbol{\theta}|\mathbf{x},d) = \prod_{i=1}^{m} l_i(\boldsymbol{\theta}_i|\mathbf{t}_i(\mathbf{x}), d)$$

where $l_i(\boldsymbol{\theta}_i|\mathbf{t}_i(\mathbf{x}))$ is fn.of $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}_i$ and $\mathbf{t}_i(\mathbf{x})$ is a function of the data $\mathbf{x}$, $i = 1, 2, 3, \ldots, m$, for each $d \in D$.

Definition

A collective is *panel independence* (pi) at time t iff it believes $\amalg_{i=1}^{m} \boldsymbol{\theta}_i$ given any $d \in D$.

Examples of Panel Independence in Probabilistic Collectives

- BNs: Panels donate distribution of parameters of a variable given its parents. Panel independence $\backsim$ global independence.
- Context specific or object orientated BNs. Single panels need to be jointly responsible for shared cpts.
- Chain graphs: One panel responsible for each each box of variables conditional on parents.
- MDM structures (Queen and Smith, 1993) Panels donate dynamic regression states.
- CEG. Smith(2010) example cites Panels donate parts of the tree: juror, forensic scientist, court and judicial statistician.

And so on...

Panel independence, Panel Separability and Distributivity

Density $\pi(\boldsymbol{\theta})$ over $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$, both collectively and individually $\pi(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_i(\boldsymbol{\theta}_i)$.

1. Panel G_i updates prior $\pi_i(\boldsymbol{\theta}_i)$ only with function $\mathbf{t}_i(\mathbf{x}_t)$ of $\mathbf{x}_t$. to obtain posterior $\pi_i^{(t)}(\boldsymbol{\theta}_i) \propto l_i(\boldsymbol{\theta}_i|\mathbf{t}_i(\mathbf{x}_t))\pi_i(\boldsymbol{\theta}_i)$, $i = 1, \ldots, m$.
2. Prior panel independence $\Rightarrow \pi^{(t)}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_i^{(t)}(\boldsymbol{\theta}_i)$.
3. EB preserved wrt separable likelihoods. If panels use the log pool to combine judgments then the collective is also EB with respect to all the *individual* experts and their panel margins.
4. But what protocols are most informative to which situations.

Ordering Experiments using Strong Likelihood Principle

Key idea: Only update on functions of data whose associated likelihood separates!

Definition

Experiments $\mathcal{E}_1$ with likelihood $l_1(\boldsymbol{\theta}|\mathbf{x})$ and $\mathcal{E}_2$ with likelihood $l_2(\boldsymbol{\theta}|\mathbf{x}')$ are equivalent (written $\mathcal{E}_1 \backsim \mathcal{E}_2$) for $\boldsymbol{\theta}$ if for all possible values of $\mathbf{x}$, and for some maps $\boldsymbol{\tau} : \mathcal{X} \rightarrow \mathcal{X}'$, $\mathbf{x} \longmapsto \boldsymbol{\tau}(\mathbf{x}) = \mathbf{x}'$ and $\boldsymbol{\tau}' : \mathcal{X} \rightarrow \mathcal{X}'$, $\mathbf{x}' \longmapsto \boldsymbol{\tau}'(\mathbf{x}') = \mathbf{x}$

$$l_2(\boldsymbol{\theta}|\boldsymbol{\tau}(\mathbf{x})) = l_1(\boldsymbol{\theta}|\mathbf{x}) \text{ and } l_1(\boldsymbol{\theta}|\boldsymbol{\tau}'(\mathbf{x}')) = l_2(\boldsymbol{\theta}|\mathbf{x}')$$

Ordering Experiments and Redundancy

Definition

Say $\mathcal{E}_1$ is *dominated* by $\mathcal{E}_2$ (written $\mathcal{E}_1 \preceq \mathcal{E}_2$) for $\boldsymbol{\theta}$ if $\exists$ experiments $\widetilde{\mathcal{E}}_2(\mathbf{x}) \backsim \mathcal{E}_1(\mathbf{t}(\mathbf{x}))$ and experiments $\widetilde{\mathcal{E}}_2(\mathbf{x}) \sim \mathcal{E}_2(\mathbf{x})$ s.t. $\widetilde{\mathcal{E}}_2(\mathbf{x})$ consists of $\widetilde{\mathcal{E}}_1(\mathbf{t}(\mathbf{x}))$ and then subsequently observing more units and/or taking additional observations whose distribution - extra $\mathcal{E}_{2:1}(\mathbf{x}|\mathbf{t}(\mathbf{x}))$ - whose associated distribution also depends only on $\boldsymbol{\theta}$. Write $\mathcal{E}_1 \prec \mathcal{E}_2$ if $\mathcal{E}_1 \preceq \mathcal{E}_2$ and $\mathcal{E}_1 \nsim \mathcal{E}_2$.

If $\mathcal{E}_i$ has likelihood $l_i(\boldsymbol{\theta}|\mathbf{x})$ $i = 1, 2$ and $\mathcal{E}_1 \preceq \mathcal{E}_2$

$$l_2(\boldsymbol{\theta}|\mathbf{x}) = l_1(\boldsymbol{\theta}|\mathbf{t}(\mathbf{x}))l_{2:1}(\boldsymbol{\theta}|\mathbf{x})$$

where $l_{2:1}(\boldsymbol{\theta}|\mathbf{x}) \propto p_{2:1}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{t}(\mathbf{x})))$ the sample density of data from the additional experiment $\mathcal{E}_{2:1}(\mathbf{x}|\mathbf{t}(\mathbf{x}))$.

Cores of experiments

Definition

Experiment $\mathcal{E}^*$ is a *core* of $\mathcal{E}$ iff $\mathcal{E}$ is panel separable, $\mathcal{E}^* \preceq \mathcal{E}$ and there is no other separable experiment $\mathcal{E}'$ s. t. $\mathcal{E}^* \prec \mathcal{E}' \preceq \mathcal{E}$

- When $\mathcal{E}$ is separable it is equal to its core.
- Sometimes a protocol needs to establish which core to choose.
- If $\mathcal{E}$ not separable then it has a subexperiment that is.

Theorem

The combination of two independent panel separable experiments $\mathcal{E}_1$ and $\mathcal{E}_2$ is panel separable. The core of two independent panel separable experiments is contained in a combination of individual cores.

Qualitative v. Quantitative sensitivities

Theorem

Suppose $\mathcal{E}_1$ - n random discrete measurements of n units $\mathbf{x}$ *has mass function*

$$p(\mathbf{x}|\boldsymbol{\theta}) = c(\mathbf{x}) \prod_{i=1}^{m} p_i(t_i(\mathbf{x})|\boldsymbol{\theta}_i, f_i(t^{(i-1)}(\mathbf{x})))$$

where $f_i(t^{(i-1)}(\mathbf{x}))$ *fn. of* $\mathbf{x}$ *only through* $(t_1(\mathbf{x}), t_2(\mathbf{x}), \ldots, t_{i-1}(\mathbf{x}))$, $p_i(t_i(\mathbf{x})|\boldsymbol{\theta}, f_i(t^{(i-1)}(\mathbf{x})))$ *fn. only of its arguments, and* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$ *takes values in* $\Theta = \Theta_1 \times \Theta_2 \times \ldots \times \Theta_m$. *Then* $\mathcal{E}_1 \sim \mathcal{E}_2$ *of m sets of stratified random samples. The first set corresponds to taking a random sample of n units where we observe the same values* $t_1(\mathbf{x})$ *as we did in* $\mathcal{E}_1(\mathbf{x})$. *For the* i^{th} *set of randomised experiments* $i = 2, \ldots, m$ *are stratified according to the levels of their conditioning set.*

Thus sample each level of $f_i(t^{(i-1)}(\mathbf{x})) \; \# \left\{ f_i(t^{(i-1)}(\mathbf{x})) \right\}$ times,

Causality and designed experiments

Experimental information can also be used by the panels. But then need additional causal assumptions.

Theorem

When the collective agrees that G is a causal Bayesian Network and parameters of different variables in the system respect global independence. at any time t: then system remains distributed under a likelihood composed of ancestral sampling experiments.

An observational data set to update.

$$\otimes \longrightarrow \odot \longrightarrow \otimes \longrightarrow \otimes$$

Discussion

- Distributive Networks surprisingly easy to build and form a fruitful and useful area of theoretical development.
- Panel independence critical! Admissibility of data critical!
- Directional conditioning of panels almost essential for distributivity.
- Approximations or simply valid partial inference?.
- Often,form of utility function,only requires panels to donate a few moments (e.g. see Queen example). When this is the case modification of ideas of separability and generalisations of LB (Goldstein and Wooff) simplifies. Collective a partial Bayesian? Panels also partial Bayesians
- Because outputs are ofen polynomial these ammenable to study through algebraic geometry.

Thank you Thank you Thank you

THANK YOU FOR YOUR ATTENTION!!!

My References

Smith J.Q.and Zwiernich P.(2011) "Bayesian Inference in Distributed Expert Systems" CRiSM report (in prep)
Smith J.Q. and Zwiernich P. (2011) " The geometry of networked Bayesian Decision Support" CRiSM Report (in prep)
Freeman, G. and Smith, J.Q. (2011) " Bayesian MAP Selection of Chain Event graphs" JMVA (to appear)
Smith, J.Q. (2010) "Bayesian Decision Analysis" Cambridge University Press
Daneshkhah, A. & Smith, J.Q.(2004) "Multicausal prior families, Randomisation and Essential Graphs" Adv. in BNs, 1-17
Faria, A.E. and Smith, J.Q. (1997) "Conditionally externally Bayesian pooling operators in chain graphs", Ann Stats. 25,4,1740-1761
Smith, J.Q.et al (1997). "Probabilistic Data Assimilation with RODOS". Radiation Protection Dosimetry,73, 57-59.
Queen, C.M., and Smith, J.Q. (1993). "Multi-regression dynamic models". J.R. Statist.Soc.B, 55, 4,849-870.

A few other references

Clemen, R.T. and Winkler , R.L (2007) "Aggregating Probability Distributions" .in Advances in Decision Analysis Cambridge University Press 154 -176

Goldstein, M. and Wooff, D. (2007) "Bayesian Linear Statistic: Theory and Methods" Wiley

Xiang, Y.(2002) "Probabilistic Reasoning in Multiagent Systems" Cambridge University Press

Caminada, G., French, S., Politis, K. and Smith, J.Q. (1999) “Uncertainty in RODOS” Doc. RODOS(B) RP(94) 05,.