Correlated Binary Variables and Multi-level Probability Assessments

Michail Papathomas
Imperial College London
Department of Epidemiology and Public Health
email:m.papathomas@imperial.ac.uk

December 14, 2009

A practical problem

An insurance company is interested in assessing the probability of paying out on a number of contested claims regarding fire damage.

- Consider six claims $B_1, B_2, ..., B_6$.

The company lawyer provides the following **probabilities of paying out**.

	B_1	B_2	B_3	B_4	B_5	B_6
$P(B_i)$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**

- **What about the probability that both cases 3 and 4 will be lost?**

Later, a specialized lawyer with vast experience in claims of this type provides **increased** probabilities so that $\mathbf{P(B_3) = 0.6}$ and $\mathbf{P(B_4) = 0.75}$.

- **How should these new assessments change our beliefs for the claims?**

Formulation of the problem

- Six events that cases B_i, $i = 1, 2, 3, 4, 5, 6$ will be lost.
- Correlated binary variables $y_1, y_2, y_3, y_4, y_5, y_6$.

$$P(B_i \text{ is lost}) = P(y_i = 1)$$

- Three levels of uncertainty

The experts provide with probabilities $p_{i,j}$ that the events are realized, $1 \leq i \leq 6$, $0 \leq j \leq 2$.

Level zero	y_1	y_2	y_3	y_4	y_5	y_6
Level one	$p_{1,1}$	$p_{2,1}$	$p_{3,1}$	$p_{4,1}$	$p_{5,1}$	$p_{6,1}$
Level two	$p_{1,2}$	$p_{2,2}$	$p_{3,2}$	$p_{4,2}$	$p_{5,2}$	$p_{6,2}$

For example,

Level zero	y_1	y_2	y_3	y_4	y_5	y_6
Level one	$p_{1,1}$	$p_{2,1}$	$\mathbf{0.6}$	$\mathbf{0.75}$	$p_{5,1}$	$p_{6,1}$
Level two	$\mathbf{0.1}$	$\mathbf{0.2}$	$\mathbf{0.3}$	$\mathbf{0.4}$	$\mathbf{0.5}$	$\mathbf{0.6}$

General formulation

- Events E_i, $i = 1, ..., n$
- Correlated binary variables $y_1, y_2, ..., y_n$

$$P(E_i \text{ is realized}) = P(y_i = 1)$$

- (J+1) levels of uncertainty

The experts provide with probabilities $p_{i,j}$, $i = 1, ..., n$, $j = 0, 1, ..., J$, that the events are realized.

Level zero	y_1	y_2	$\cdots$	y_n
Level one	$p_{1,1}$	$p_{2,1}$	$\cdots$	$p_{n,1}$
Level two	$p_{1,2}$	$p_{2,2}$	$\cdots$	$p_{n,2}$
$\vdots$				
Level J	$p_{1,J}$	$p_{2,J}$	$\cdots$	$p_{n,J}$

The model

The joint density between the binary variables and their assessments is defined as,

$$P(\boldsymbol{y}, \boldsymbol{p}_1 | \boldsymbol{p}_2) = P(\boldsymbol{y} | \boldsymbol{p}_1, \boldsymbol{p}_2) \times P(\boldsymbol{p}_1 | \boldsymbol{p}_2).$$

- $\boldsymbol{p}_j$ denotes all probability assessments at level j.

Constructing $P(\boldsymbol{y}|\boldsymbol{p}_1, \boldsymbol{p}_2)$ (threshold copula)

For assessments at level one, we relate the n binary variables to a **latent** vector $\mathbf{x}$ (Papathomas and O'Hagan, 2005), so that,

$$\mathbf{x} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta_{1,2}^{(1)} & \cdots & \delta_{1,n}^{(1)} \\ \delta_{2,1}^{(1)} & 1 & & \delta_{2,n}^{(1)} \\ \vdots & & & \\ \delta_{n,1}^{(1)} & \delta_{n,2}^{(1)} & & 1 \end{pmatrix}\right),$$

where index (1) represents the level where assessments are considered. Then,

$$P(y_1 = 1, ..., y_n = 0) = P(x_1 < \Phi^{-1}(p_{1,1}), ..., x_n \geq \Phi^{-1}(p_{n,1})).$$

- The $\delta_{1,2}^{(1)}$ correlation, for instance, is elicited by asking for probabilities $P(y_1 = 1, y_2 = 1|p_{1,1}, p_{2,1})$ or $P(y_1 = 1|y_2 = 1, p_{1,1}, p_{2,1})$

Constructing $P(\boldsymbol{p}_1|\boldsymbol{p}_2)$

- We consider the following underlying structure.

$$\begin{array}{lllll} z_1 & z_2 & z_3 & z_4 & z_5 \\ & & & & & \}\mathbf{E}_1 = (E_{1,1}, E_{2,1}, E_{3,1}, E_{4,1}, E_{5,1}) \\ z_{1,1} & z_{2,1} & z_{3,1} & z_{4,1} & z_{5,1} \\ & & & & & \}\mathbf{E}_2 = (E_{1,2}, E_{2,2}, E_{3,2}, E_{4,2}, E_{5,2}) \\ z_{1,2} & z_{2,2} & z_{3,2} & z_{4,2} & z_{5,2} \end{array}$$

- $\mathbf{E}_j$ normal, independent with mean zero.
- $\text{Var}(\mathbf{E}_2) = \boldsymbol{\Sigma}_{\boldsymbol{z}_{1,2}}$ and therefore $\boldsymbol{z}_1|\boldsymbol{z}_2 \sim N(\boldsymbol{z}_2, \boldsymbol{\Sigma}_{\boldsymbol{z}_{1,2}})$
- We parametrise so that the diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{z}_{1,2}}$ are $v_{i,2}^2 - v_{i,1}^2$

Constructing $P(\boldsymbol{p}_1|\boldsymbol{p}_2)$ (continued)

For each $p_{i,j}$ **at level greater than zero**, consider a $z_{i,j}$ so that

$$z_{i,j} = -v_{i,j}\Phi^{-1}(p_{i,j}).$$

Then,

$$P(\boldsymbol{p}_1|\boldsymbol{p}_2) = P(\boldsymbol{z}_1|\boldsymbol{z}_2)|(\frac{d\boldsymbol{z}_1}{d\boldsymbol{p}_1})|$$

$$= \{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{z_{1,2}}})\}^{-\frac{1}{2}} \times \exp\{\frac{1}{2}(\boldsymbol{V}_1\boldsymbol{\phi}_1 - \boldsymbol{V}_2\boldsymbol{\phi}_2)^\top\boldsymbol{\Sigma}_{\boldsymbol{z_{1,2}}}{}^{-1}(\boldsymbol{V}_1\boldsymbol{\phi}_1 - \boldsymbol{V}_2\boldsymbol{\phi}_2)\}$$

$$\times(2\pi)^{\frac{n}{2}}\det(\boldsymbol{V}_1)\exp\{\frac{1}{2}\boldsymbol{\phi}_1^\top\boldsymbol{\phi}_1\}.$$

- $\boldsymbol{\phi_j} = (\Phi^{-1}(p_{1,j}), ..., \Phi^{-1}(p_{n,j}))^\top$
- $\boldsymbol{V}_j = \text{diag}\{v_{1,j}, v_{2,j}, ..., v_{n,j}\}$

The marginal density of an assessment

The density $P(p_{i,1}|p_{i,2})$ is controlled by the ratio $v_{i,1}/v_{i,2}$. It represents the decision makers opinion on **the amount of learning that is taking place** when obtaining assessments at level 1.

- $\mathrm{E}(p_{i,1}|p_{i,2}) = p_{i,2}$.
- $$\mathrm{E}(p_{i,1}^2|p_{i,2}) = P(\xi_1 < \Phi^{-1}(p_{i,2}), \xi_2 < \Phi^{-1}(p_{i,2})) \qquad (1)$$
  where,

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 - \frac{v_{i,1}^2}{v_{i,2}^2} \\ 1 - \frac{v_{i,1}^2}{v_{i,2}^2} & 1 \end{pmatrix} \right).$$

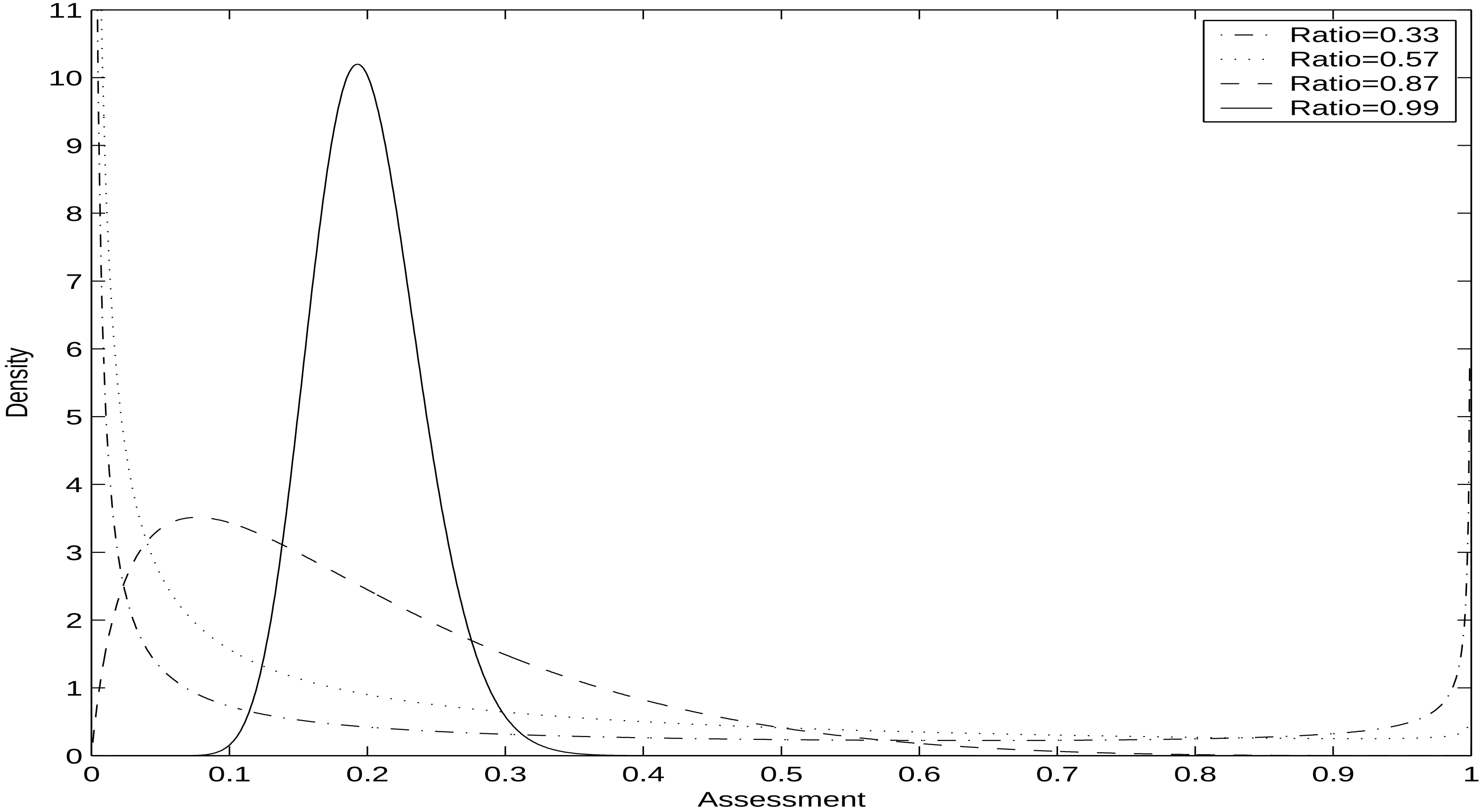

Figure 1: Density of $p_{i,1}$ for $p_{i,2} = 0.2$ and various ratios $v_{i,1}/v_{i,2}$

Eliciting the $v_{i,j}$ parameters

The **expected variance reduction** for y_i, after a future $p_{i,1}$ is received is,

$$\mathrm{Var}(y_i|p_{i,2}) - \mathrm{E}_{p_{i,1}}\{\mathrm{Var}(y_i|p_{i,1},p_{i,2})\} = \mathrm{E}(p_{i,1}^2|p_{i,2}) - p_{i,2}^2.$$

If the expert believes that from level of uncertainty 2 to level 1 there is an Ω per cent expected reduction in the variance of y_i then,

$$\mathrm{E}(p_{i,1}^2|p_{i,2}) - p_{i,2}^2 = \frac{\Omega}{100}\{p_{i,2}(1-p_{i,2})\} \qquad (2)$$

Steps

- Set an arbitrary value for $v_{i,2}$, say $v_{i,2} = 1$.
- Elicit the expert's or decision maker's expectation on the proportional reduction of uncertainty Ω for y_i when an assessment is obtained at level 1.
- This expectation gives the $v_{i,1}$ parameter.

Eliciting correlations for the $\Sigma_{z_{1,2}}$ matrix

- To complete the elicitation of $\mathbf{\Sigma}_{\boldsymbol{z}_{1,2}}$ we need to elicit its correlations
- To simplify this elicitation, we impose that **the fully specified joint density should be consistent with the threshold copula at level 2.**

Notice that,

$$P(\boldsymbol{y}|\boldsymbol{p}_2) = \int_0^1 \dots \int_0^1 P(\boldsymbol{y}|\boldsymbol{p}_1, \boldsymbol{p}_2) \times P(\boldsymbol{p}_1|\boldsymbol{p}_2) d\boldsymbol{p}_1.$$

- Therefore, the above equality should hold with the probabilities $P(\boldsymbol{y}|\boldsymbol{p}_1, \boldsymbol{p}_2)$ and $P(\boldsymbol{y}|\boldsymbol{p}_2)$ obtained using the threshold copula at levels 1 and 2 respectively.

Theorem : The joint density $P(\boldsymbol{y}, \boldsymbol{p}_1 | \boldsymbol{p}_2)$ is consistent with the threshold copula if,

$$\boldsymbol{\Sigma}_{\boldsymbol{z_{1,2}}} = \boldsymbol{V}_2 \boldsymbol{\Sigma}_2 \boldsymbol{V}_2 - \boldsymbol{V}_1 \boldsymbol{\Sigma}_1 \boldsymbol{V}_1 \qquad (3)$$

(Papathomas, 2008, Scandinavian Journal of Statistics)

A practical problem (continued)

Consider 30 claims divided in two strata $\{y_1, ..., y_{15}\}$, $\{y_{16}, ..., y_{30}\}$.

Assume the following elicited copula parameters **for level 2**:

- $\delta_2^s = 0.45$ (between claims that belong to the same stratum)
- $\delta_2 = 0.2$ (between claims from different strata)

Assume the following elicited copula parameters **for level 1**:

- $\delta_1^s = 0.40$
- $\delta_1 = 0.15$.

For both strata, and a claim with a 0.5 prior probability, the expert expects **a 60% reduction in uncertainty** when assessments are obtained at level 1. For $v_2 = 1$, this percentage translates to $v_1 = 0.443$.

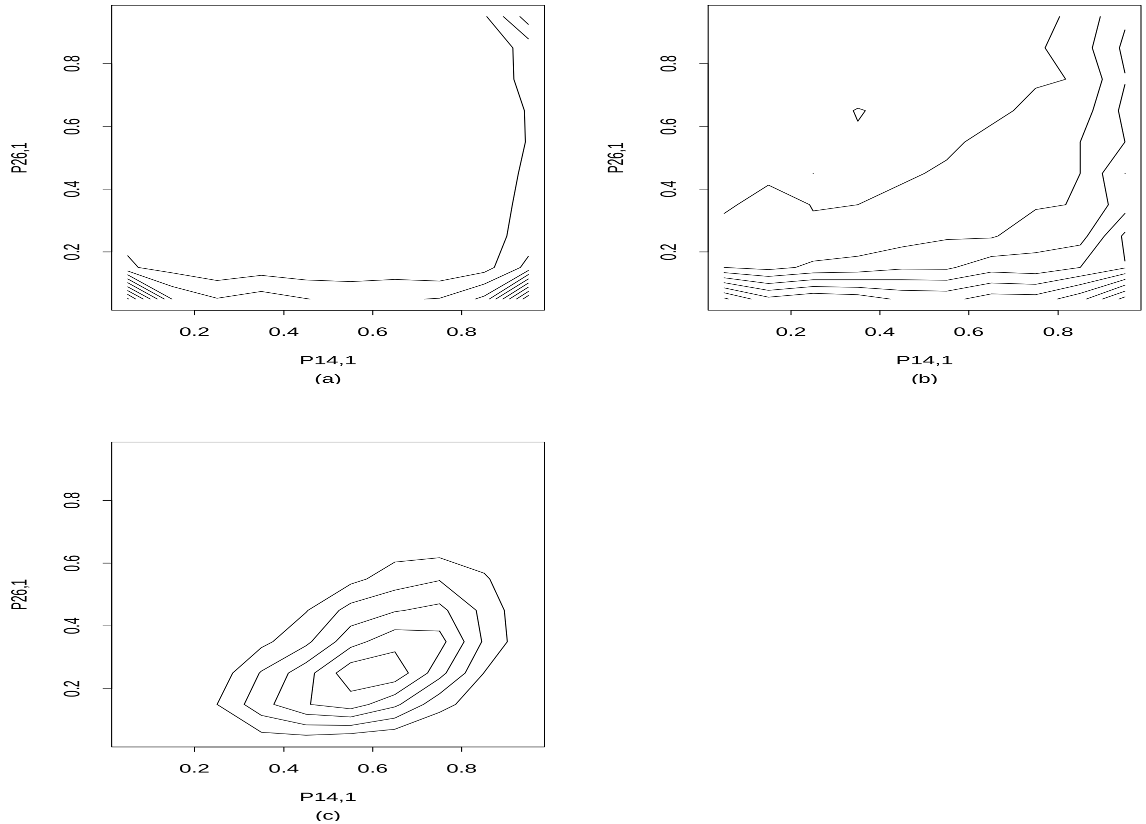

Figure 2: Contour plots of the joint density of $p_{14,1}$ and $p_{21,1}$ for $p_{14,2} = 0.6$ and $p_{26,2} = 0.3$: (a) 60% expected reduction of uncertainty; (b) 40% expected reduction of uncertainty; (c) 10% expected reduction of uncertainty

Table 1. Elicited marginal probabilities and posterior marginal probabilities when assessments at levels 2 and 1 are considered and when assessments at levels 2,1 and 0 are considered

	First Stratum						Second Stratum				
i	Lev 2	Lev 1	Post.	Observed	Post.	i	Lev 2	Lev 1	Post.	Observed	Post.
1	0.8	-	0.9073	-	0.9281	16	0.15	-	0.1128	-	0.1164
2	0.9	-	0.9659	-	0.9739	17	0.6	0.2	0.2086	-	0.2103
3	0.1	0.25	0.2518	-	0.2849	18	0.9	-	0.9351	-	0.9315
4	0.1	-	0.1136	-	0.1281	19	0.35	-	0.3168	-	0.3286
5	0.2	-	0.2512	-	0.2666	20	0.35	0.08	0.0807	-	0.0898
6	0.35	-	0.4537	-	0.474	21	0.2	-	0.1671	-	0.1605
7	0.7	0.8	0.7917	1	1	22	0.65	-	0.6706	-	0.682
8	0.7	-	0.833	-	0.8476	23	0.8	0.9	0.8963	-	0.9115
9	0.15	-	0.1869	-	0.1999	24	0.7	-	0.7257	-	0.7444
10	0.5	-	0.6404	-	0.6501	25	0.5	-	0.4914	-	0.5062
11	0.95	-	0.9909	-	0.9917	26	0.3	0.8	0.7961	-	0.8122
12	0.55	0.9	0.898	1	1	27	0.95	-	0.9746	-	0.9754
13	0.55	-	0.6907	-	0.7133	28	0.1	-	0.0609	-	0.0674
14	0.6	0.9	0.8987	-	0.9281	29	0.8	-	0.8342	-	0.8444
15	0.6	-	0.7413		0.7704	30	0.4	-	0.3765	-	0.3882

Table 2. Joint prior and posterior probabilities

	Joint probabilities when assessments at level 2 are considered					Posterior probabilities when assessments at levels 2 and 1 are considered				
	Y_6	Y_7	Y_{12}	Y_{19}	Y_{20}	Y_6	Y_7	Y_{12}	Y_{19}	Y_{20}
Y_6	0.3479	0.2997	0.2554	0.1485	0.1535	0.4537	0.3838	0.42	0.1504	0.0411
Y_7		0.6976	0.45	0.2661	0.2728		0.7917	0.7355	0.256	0.0706
Y_{12}			0.5469	0.2187	0.2224			0.898	0.2891	0.0746
Y_{19}				0.3449	0.1911				0.3168	0.038
Y_{20}					0.3518					0.0807

	Posterior probabilities when assessments at levels 2,1 and 0 are considered				
	Y_6	Y_7	Y_{12}	Y_{19}	Y_{20}
Y_6	0.474	0.474	0.474	0.1676	0.0482
Y_7		1	1	0.3286	0.0898
Y_{12}			1	0.3286	0.0898
Y_{19}				0.3286	0.0396
Y_{20}					0.0898

A practical problem (continued)

Assume that the cost for paying out on a claim is 50.
Let $w = 50 \times \sum_1^{30} y_i$.

We can easily calculate from the simulated sample that,

$$\mathrm{E}\{w|\boldsymbol{p}_2\} = 779$$

$$\mathrm{E}\{w|\boldsymbol{p}_1(s_1), \boldsymbol{p}_2\} = 863.1$$

$$\mathrm{E}\{w|\boldsymbol{y}(s_0), \boldsymbol{p}_1(s_1), \boldsymbol{p}_2\} = 896.4$$

For the variance of w we obtain that,

$$\mathrm{Var}\{w|\boldsymbol{p}_2\} = 75,444$$

$$\mathrm{Var}\{w|\boldsymbol{p}_1(s_1), \boldsymbol{p}_2\} = 29,961$$

$$\mathrm{Var}\{w|\boldsymbol{y}(s_0), \boldsymbol{p}_1(s_1), \boldsymbol{p}_2\} = 24,473$$

Discussion

We have developed a fully specified joint density for a set of correlated binary variables with multi-level probability assessments. This density is

- Rich enough to describe a wide range of practical problems
- Lends itself to inference via simulation (MCMC). Sampling from truncated multivariate normal distributions is required. To sample efficiently we use auxiliary random vectors.
- Possible to elicit asking simple “layman’s” questions

Discussion (continued)

- To elicit a correlation for the threshold copula at, say, level 1, we can ask for a joint probability like

$$P(y_1 = 1, y_2 = 1 | p_{1,1}, p_{2,1})$$

  and/or a conditional probability

$$P(y_1 = 1 | y_2 = 1, p_{1,1}, p_{2,1}).$$

  providing probability bounds and guidance to the expert.
- These probabilities can be inconsistent, giving **considerably different values for the correlation**, depending on the marginal probabilities and conditioning.
- A careful validation procedure is necessary where plenty of feedback is provided to the expert.

Discussion (continued)

- In a recent consultation, experts were more comfortable in providing conditional probabilities

$$P(y_1 = 1 | y_2 = 1, y_3 = 1, y_4 = 0, y_5 = 0, y_6 = 0, y_7 = 0, p_{1,1}, p_{2,1}, \ldots)$$

$$P(y_1 = 1 | y_2 = 1, y_3 = 1, y_4 = 1, y_5 = 0, y_6 = 0, y_7 = 0, p_{1,1}, p_{2,1}, \ldots)$$

$$P(y_1 = 1 | y_2 = 1, y_3 = 1, y_4 = 1, y_5 = 1, y_6 = 0, y_7 = 0, p_{1,1}, p_{2,1}, \ldots)$$

- Probabilities were often inconsistent, giving **higher correlations as the number of observed $y_i = 1$ increased**.

Discussion (continued)

- In a recent consultation, experts were more comfortable in providing conditional probabilities

$$P(y_1 = 1 | y_2 = 1, y_3 = 1, y_4 = 0, y_5 = 0, y_6 = 0, y_7 = 0, p_{1,1}, p_{2,1}, \ldots)$$

$$P(y_1 = 1 | y_2 = 1, y_3 = 1, y_4 = 1, y_5 = 0, y_6 = 0, y_7 = 0, p_{1,1}, p_{2,1}, \ldots)$$

$$P(y_1 = 1 | y_2 = 1, y_3 = 1, y_4 = 1, y_5 = 1, y_6 = 0, y_7 = 0, p_{1,1}, p_{2,1}, \ldots)$$

- Probabilities were often inconsistent, giving **higher correlations as the number of observed** $y_i = 1$ **increased**.
- Sometimes expert beliefs were ‘asymmetric’ concerning the dependence between two groups of assets. For instance, the expert insisted that learning about new steel structures **should affect** beliefs for old steel structures (element of surprise). However, learning about old steel structures **should not affect** beliefs for new steel structures.

Discussion (continued)

- Expected reduction in uncertainty, in percentage terms, can be translated to $v_{i,j}$ parameters using variance, standard deviation or some other measure.
- The statistical estimation of the $v_{i,j}$ parameters could be the subject of future research, but assessments are sparsely elicited at low levels of uncertainty and this may present with difficulties.

Discussion (continued)

Assume elicited parameters so that

- $\delta_2^s = 0.15$, $\delta_2 = 0.12$
- $\delta_1^s = 0.45$, $\delta_1 = 0.3$
- $v_1 = 0.646$ (40% expected reduction of uncertainty)

Now, the probability that $y_2 = 1$ drops from 0.9 to 0.885.
The probability that $y_{14} = 1$ increases from 0.9 to 0.9352.

The apparently incoherent posterior probabilities are caused by the negative correlations between future assessments at level 1. It is easy to deduce from equation (4) that the negative correlations occurred because the copula correlations at level 1 are higher than the corresponding ones at level 2, whilst the expected reduction of uncertainty is not high enough to compensate for this difference.

Discussion (continued)

- When the variance of the probability assessments is under-estimated, elicited assessments close to 0 or 1 become influential outliers and have a strong effect on the posterior probabilities.

  Possible remedies:

  - Implementing the model with software that will produce warning flags.
  - The true 0 or 1 value is assigned at level zero, even if the relevant information corresponds to a higher level of uncertainty.

References

Goldstein, M. and O'Hagan, A. (1996): Bayes linear sufficiency and systems of expert posterior assessments. *J. R. Statist. Soc. B*, vol. **58**, pp. 301-316.

Jeffrey, R.C. (1983): *The logic of decision. 2nd Ed.*, The University of Chicago Press, London.

Papathomas, M. and Hocking, R.J. (2003): Bayesian updating for binary variables: an application in the UK water industry. *The Statistician*, **52**, 483-499.

Papathomas, M. and O'Hagan, A. (2005): Updating beliefs for binary variables. *Journal of Statistical Planning and Inference.* **135**, pp. 324-338.

Papathomas, M. (2008): A fully parametric model for correlated binary variables with multi-level probability assessments *Scandinavian Journal of Statistics.* **35**, pp. 169-185.