Bayes Linear Methodology and Extracting information from experts for use of Bayes Linear Analysis.

V. Vijfvinkel

Contents

1 Introduction

This paper discusses the Bayes Linear Methodology (BLM) and the use of *elicitation* of experts in this context. When analyzing a system with an accurate data collection and well defined relations, a statistician can accurately give information about the expectation of different components. When data and relations are not easily gathered, like in processes that only come by rarely or processes with complex structures, other methods of statistical analysis have to be explored. Especially in case of decision making it is not easy to value the expected payback.

When time and money are not important, extensive and time consuming models can be used to model the process and these can help in making decisions. But unlimited time and funding is never really the case so choices about the type of approach have to be taken. In this paper, a model is discussed that can calculate the expected output of the decisions at hand, with few exact details about the underlying distributions of the variables of influence.

A model for a process like decision making can not be made without interviewing specialists in fields that cover different parts of the process. The extraction of information from these experts is called *elicitation*. Because experts on the subject are rarely also statistician a procedure has to be started between the people who know the subject and people who know how to value this knowledge.

In the first section of this paper the fundaments of BLM are discussed. The following section gives a description of problems with experts elicitation and gives a framework for constructive use of expert judgement. This section is followed by three suggested methods of elicitation. The fifth section describes an approach to elicitation when the experts differ in opinion. This paper is concluded with a summary of the topics.

2 Bayes Linear Methodology

The Bayes Linear Methodology approach provides a tool to express beliefs and to combine judgement on uncertainty with observational data. Contrary to Bayes Analysis, BLM does not use probability distributions for expressing these judgements. These judgements are expressed by their expectations. This is especially practical when dealing with large complex problems, where specifying a full joint probability distribution would be a difficult task because there are a lot of factors that influence the problem. Using the expectation directly as the basic quantity and specifying the expectation rather than a full probability distribution has been the topic of several studies. A term often used for an expectation elicited directly is *prevision*, introduced by Finetti, see [6]. Also various operational definitions for prevision are proposed by Finetti. Because BLM is often applied to very complex problems, choices have to be made about which aspect of our prior belief we want to quantify and to which detail we want to do this. These choices depend on several aspects of the problem:

- the degree of interest in the different parts of the problem
- the ability to specify each aspect of uncertainty
- the amount of time and effort that can be spent on the problem
- the detail of prior specification that is necessary in order to extract the information from the data

2.1 Why BLM

Bayesian theory can be of great value to a decision making model. The modelling of a full probabilistic model however, can be a time consuming project. The elicitation of quantities alone is hard because of the required detail of a full prior distribution. When time is not sufficiently available or certainty of the expert about details of the subject is small, the precise modelling will not automatically lead to a model of which the output can be fully trusted. When there is only certainty about the broad conclusion of the output, the detailed elicitation of the experts seems superfluous. But the highly detailed model is necessary for a full Bayesian methodology. BLM gives an opportunity to make a model with less detailed elicitation and modelling. By only using the mean, variance and covariance as input to the model, the process of elicitation is simplified. Also the complex approach of full Bayesian theory is avoided by using BLM.

2.2 The base of the BLM

In order to express the judgements the following model is used:
$C = (X_1, X_2, X_3, \ldots, X_n)$ is a collection of ordered random variables. These random variables are the variables about which statements of uncertainty are made. The collection C is called the base. For each pair (X_i, X_j) , with $X_i, X_j \epsilon C$ the following will be specified:

1. the expectation, $E(X_i)$, giving a simple quantification of the belief as to the magnitude of X_i
2. the variance, $Var(X_i)$, quantifying the uncertainty in the judgements of the magnitude of X_i (all elements X_i have a finite prior variance)
3. the covariance, $Cov(X_i, X_j)$, expressing a judgement on the relationship between the quantities, quantifying the extent to which observations on X_j may influence the belief as to the size of X_i

The size of C determines the detail of the analysis; by extending C it is possible to obtain the full joint probability distribution of the elements of C. It is appropriate for large problems to restrict the investigation to a collection that represents the prior belief structure. The formal structure in which the beliefs are represented is formed with C. With the collection C the linear space $\langle C \rangle$ is constructed. This space consists of all finite linear combinations

$$h_0 X_0 + h_1 X_{i1} + \ldots \ + h_k X_{ik}$$

of the elements of C, where X_0 is the unit constant and X_i is a subset of C with k elements, with X_{in} being the *n-th* element of X_i The space $\langle C \rangle$ can be viewed as a vector space in which each X_i is a vector and linear combinations of vectors are the corresponding linear combinations of the random quantities. Covariance defines an inner product $(.,.)$ and norm over $\langle C \rangle$, defined, for X, Y in $\langle C \rangle$ to be

$$(X, Y) = Cov(X, Y), ||X||^2 = Var(X)$$

The vector space $\langle C \rangle$, with the covariance inner product (.,.), defines an inner product space $[C]$. This $[C]$ is called the *belief structure* with base C. Within $[C]$, the length of any vector is equal to the standard deviation of the random quantities. This space, $[C]$, gives a minimal structure for the specifications of the beliefs that suffices for general analyses. The belief structure gives the possibility to restrict the specification to any linear subspace of this structure, so that only aspects of the beliefs have to be specified that can be measured and that have a value to the research. With the given belief structure it is possible to adjust the beliefs given observed data. All expectations, variances and covariances can be adjusted after observations of data. These methods are relatively easy to comprehend and to implement. The formal properties of this approach follow from the linearity of the inner product structure. This is the reason why the term Bayes Linear is used.

2.3 Adjusting beliefs by data

After the collection C has been identified and the prior mean, variance and covariance have been specified, the prior beliefs can be adjusted using observed data. Say that a subset of C, $D = (D_1, D_2, ..., D_k)$ of observed random quantities. With this realisation the prior beliefs about a subset of C, $B = (B_1, B_2, ..., B_m)$, can be changed to also contain this new information.

2.3.1 Adjusting expectation

The expectation of the random quantity $X \in B$ after observing D, denoted as $E_D(X)$, is given by the linear combination

$$E_D(X) = h_D^T D = \sum_{i=0}^{k} h_i D_i,$$

with h_i resulting from the minimization of

$$E((X - \sum_{i=0}^{k} h_i D_i)^2)$$

where $D_0 = 1$.

The adjusted expectation is determined by the prior mean, variance and covariance. If $Var(D)$ is full of rank then:

$$E_D(X) = E(X) + Cov(X, D)[Var(D)]^{-1}(D - E(D)) \quad (*)$$

This expression for $E_D(X)$ can be attained with the values of h_i by minimization of the expression above, and then substituting h_i with these values. For illustration here the proof is given for equality $(*)$ for one observation.

Proof.

$E_D(X) = \min_{a,b} E((X - (a + bD))^2)$

$E((X - (a + bD))^2) = E[(X - a - bD)(X - a - bD))]$

$$\begin{aligned}
&= E[X^2 - 2aX - 2bXD + 2abD + b^2D^2 + a^2] \\
&= E(X^2) - 2aE(X) - 2bE(XD) + 2abE(D) + b^2E(D^2) + a^2 \\
&= a^2 + (-2E(X) + 2bE(D))a + [E(X^2) - 2bE(XD) + b^2E(D^2)] \\
&= E(D^2)b^2 + (-2E(XD) + 2aE(D))b + [E(X^2) - 2aE(X) + a^2]
\end{aligned}$$

The two last expressions are both convex in a and b so their minimum can easily be computed . Recal that equation $qx^2 + rx + s$ finds its minimum in $\frac{-r}{2q}$.This leads to the following:

$$a = \frac{-(2E(X)+2bE(D))}{2} = E(X) - bE(D)$$

$$b = \frac{E(XD)-aE(D)}{E(D^2)} \quad (**)$$

Now the values of a and b can be extracted. First the value for b is substituted. In the following the definitions $Var(D) = E(D^2) - E(D)^2$ and $Cov(X, D) = E(XD) - E(X)E(D)$ are used.

$$a = E(X) - \frac{E(D)(E(XD)-a(E(D)}{E(D^2)}$$

$$a = E(X) - \frac{E(D)E(XD)}{E(D^2)} + \frac{a[E(D)]^2}{E(D^2)}$$

$$= \frac{E(X)}{1-\frac{E(D)^2}{E(D^2)}} - \frac{E(D)E(XD)}{(1-\frac{E(D)^2}{E(D^2)})E(D^2)}$$

$$= \frac{E(X)}{\frac{E(D^2)-E(D)^2}{E(D^2)}} - \frac{E(D)E(XD)}{E(D^2)-E(D)^2}$$

$$= \frac{E(X)E(D^2)-E(D)E(XD)}{Var(D)}$$

$$= \frac{E(X)[Var(D)+E(D)^2]-E(D)E(XD)}{Var(D)}$$

$$= E(X) + \frac{E(D)^2E(X)-E(D)E(XD)}{Var(D)}$$

$$= E(X) + \frac{-E(D)(-E(X)E(D)+E(XD)}{Var(D)}$$

$$= E(X) - E(D)\frac{Cov(X,D)}{Var(D)}$$

Now a sustitute in $(**)$.

$$b = \frac{E(XD)-(E(X)-bE(D))E(D)}{E(D^2)}$$

$$b = \frac{E(XD)-E(X)E(D)}{E(D^2)} + b\frac{E(D)^2}{E(D^2)}$$

$$E(D^2) - E(D)^2 b = Cov(X,D)$$

$$\frac{Var(D)}{E(D2)}b = \frac{Cov(X,D)}{E(D^2)}$$

$$b = \frac{Cov(X,D)}{Var(D)}$$

Hence the minimum of $a + bD$ is given by

$$a + bD = E(X) - \frac{Cov(X,D)}{Var(D)}E(D) + \frac{Cov(X,D)}{Var(d)}D$$

$$= E(X) + \frac{Cov(X,D)}{Var(D)}[D - E(D)]. \qquad \square$$

This adjusted expectation obeys the following properties for any X, X_1, X_2 and constants c, d:

$$E_D(cX_1 + dX_2) = cE_D(X_1) + dE_D(X_2)$$

$$E(E_D(X)) = E(X)$$

2.3.2 Adjusting variance

Define $[X/D]$, the adjusted version of X given D, as

$$[X/D] = X - E_D(X)$$

this is also called the 'residual ' form of X. This quantity has the following properties:

1. $E([X/D]) = 0$
2. $Cov([X/D], E_D(X)) = 0$

Now X is written as

$$X = [X/D] + E_D(X)$$

The variance of X, $Var(X)$, can be split

$$Var(X) = Var([X/D]) + Var(E_D(X)).$$

The variance here is written as the sum of the variance of the expectation plus the variance. $Var(E_D(X))$ is exactly the reduction of the variance before observing D. Therefore, the variance of X after observing D is equal to $Var([X/D])$. The adjusted variance of X follows from this

$$Var_D(X) = Var([X/D]) = E((X - E_D(X))^2)$$

In relation to the prior mean, variance and covariance, $Var_D(X)$ is represented as

$$Var_D(X) = Var(X) - Cov(X, D)[Var(D)]^{-1}Cov(D, X)$$

Exact proofs of the equalities in this section can be found in [1].

3 Elicitation and expert judgement

Elicitation is the process of capturing knowledge and beliefs about uncertain quantities and form a probability distribution for those quantities. Often elicitation is concerned with formulating a probability distribution for quantities when there is no data. This situation arises in decision making, where uncertainty needs to be expressed as a probability distribution in order to derive expected utility. Elicitation can be viewed as a process between a facilitator who assists the expert in formulating the knowledge in probabilistic form. The facilitator acts on behalf of the decision maker, who after the facilitator summarized the results, makes a decision based on these results. Within the Bayes statistics the knowledge of the expert can be seen as the prior knowledge.

The use of this knowledge is not new. Halfway into the previous century the knowledge of experts was also seen as a valuable collection of information. About that period two methods were developed for mobilizing this information. Now known as classical approaches to expert analysis, the scenario analysis and the Delphi method stood their ground for two decades; a small description is given.

Scenario analysis. A scenario can be defined as a hypothetical sequence of events constructed for the purpose of focusing attention on causal processes and decision points. In scenario analysis, trends and predictable influences are used to extrapolate a surprise-free scenario. This scenario serves as a base for creating alternative situations by varying different aspects in this scenario. In this manner it is possible to tell how exactly a hypothetical situation came about. It is also possible to identify actors and the steps they can make to prevent, divert or facilitate the process. The problem here is that the result is not a probability distribution, but an analysis of the different situations and probability theory is not a part of the analysis.

The Delphi method. In this method a monitoring team defines a set of issues and selects a set of respondents who are experts on the issues. In general the respondents do not know who the other respondents are, and the responses are anonymous. A preliminary questionnaire is sent to the respondents for comments. These are then used to establish a definitive questionnaire. This questionnaire is then sent to the respondents and their answers are analyzed by the monitoring team. The set of responses is then sent back to the respondents, together with the median answer and the interquartile range, the range containing all but the lower 25% and the upper 25% of the responses. Now the experts are asked if they want to revise their initial predictions. When a respondents predictions remain outside of the interquartile range he is asked to give arguments for the predictions. The revised predictions are then processed in the same way as in the first responses, and arguments for outliers are summarized. This process is iterated typically three or four times. The responses on the final round generally show a smaller spread than those of the first round, and this is taken to indicate that the experts have reached a degree of consensus. The median values on the final round are taken as the best predictions.

3.1 Problems with elicitation

When eliciting probabilities, one has to take in account how a person assesses the probability of an event. In the field of psychology there has been a great deal of interest in elicitation. It appears that judgements of persons highly depend on only a few mental operations, these are called heuristics. Heuristics can provide somewhat effective results but they can also lead to severe errors. In the following three of these heuristics will be discussed.

Judgement by representativeness. Several errors can be made when eliciting a conditional probability, for example: what is the probability event X will generate an event Y ($P(Y|X)$). People intuitively compare the features of X and Y and then they will let the probability depend on the similarity between these to events. An error that often appears in this kind of elicitation is that the unconditional probability of Y is not taken into account. For example, if people are asked after given a description of a personality, what a more likely profession is for this person, people tend to only compare the character to the job and look for the most similarities. What is not taken into account is the fraction of people that have this profession.

Judgement by availability. When persons are elicited to estimate the frequency of an event, the ease with which they can recall an event influences the frequency or probability that they will give to an event. This is a result of the person remembering more easily the occurrences that have a greater frequency. Also events that happened recent to the elicitation could be given a greater frequency because of the fast access of this event. An example of this is the question: Is a randomly chosen English word more likely to start with an 'r' or to have an 'r' as the third letter? Because words that begin with an 'r' are remembered fast and better a greater probability will be given to these words. But there are less words that start with an 'r' than words that have an 'r' on the third place.

Judgement of anchoring and adjustment. This widely used strategy is used when a initial value (anchor) for a random quantity is given and then the person adjusts this to obtain a final estimate. The problem here is that the person is likely to make a to small adjustment. This phenomenon of making too small an adjustment is called anchoring.

Next to these intrinsic errors in the way people tend to remember things, other influences have to be taken in to account.

Law of small numbers. People tend to think that a sample of a population has the same characteristics as the population itself, even if the sample is very small.

Hindsight bias. When a person has already seen the data it is possible that he or she updates the opinion on basis of that data. When an event has already happened people tend to think it was more likely then it actually was.

Overconfidence. There is a tendency to ignore the tail of a distribution function.

For further reading about bias in opinion see [7] and [8]

3.2 Guidelines for structured elicitation

An expert's opinion can be useful in making risk analysis but it is not a source of rational consensus. Experts may differ a lot in their opinion and unstructured use of their opinion, also given the previous principles of bias of experts and the use of heuristics, is therefore not constructive in a risk analysis. A European commission set out to make a methodology, see [5], in order to get a constructive use of expert opinion. The following principles were identified:

Reproducibility. It must be possible for scientific peers to review and if necessary reproduce all calculations. This entails that the calculational models must be fully specified and the ingredient data must be made available

Accountability The source of expert subjective probabilities must be identified. Accountability to the decision maker enhances quality and credibility and must be insisted upon. The decision maker should be able to trace every subjective probability to the name of the person or institution from which it came.

Empirical control. Expert probability assessments must in principle be susceptible to empirical control. A methodology for using expert opinion should incorporate some form of empirical control, at least in principle. It must be possible to evaluate expert probabilistic opinion on the basis of possible observations.

Neutrality. The method for combining or evaluating expert opinion should encourage experts to state their true opinions. A poorly chosen method of combining/evaluating expert opinion will encourage experts to state an opinion that differs with their true opinion. A known example is the technique used by Delphi. Here the expert is "punished" when he deviates strongly from a median value and the expert is "rewarded" when the opinion is changed towards the median.

Fairness. All experts are treated equally, prior to processing the results of observations

3.3 Eliciting Variance

In most Bayesian research it is necessary to elicit variance. Experiments have shown that it is difficult for many people to understand the idea of variance and also to give an indication of this. A way of eliciting variance is to ask an interval so that direct assessment is not necessary. When several assumptions on the underlying distribution are made, this way of extracting the variance can be effective. Here two ways of extracting an interval are discussed.

Fixed interval method For a random quantity X all possible outcomes are divided by the facilitator in several disjunct intervals. The expert is asked to point out the interval that is most likely to obtain X and to assign a probability to this interval. After this the interval second most likely is asked and so on. These subjective probabilities are then transformed under the constraint that their sum must be one. Now the variance can be extracted from the data.

Variable interval method Here the expert is asked to partition the outcome space himself. By a line of questioning like: at what point is it equally probable that X is greater or less then the value of this point. In this manner the space is divided in two disjunct intervals and again the question is asked for both intervals. The advantage of this method is that the expert only deals with equal probabilities, which are easier to comprehend.

4 Elicitation of mean, variance and covariance

Although the BLM, in comparison to standard Bayes analysis, can reduce the effort needed to analyse complex situations, it still requires the analyst to make specifications of moments. Here are three possible ways of doing so:

- Parametric fitting: Assessing joint quantiles for uncertain quantities, and then fitting a distribution from some parametric family and deriving the appropriate moments.
- Constructing Bayes linear graphical models: these graphical models relate the quantities of interest to other quantities for which we may more easily specify the moments.
- Analogous data set: Consider a data set or sets that have a similar character to the variable under discussion, and subjectively determine appropriate adjustment factors for mean and variance.

In this chapter these methodologies will be discussed. Next to this method another approach on finding mean and variance will be brought to attention. Parametric fitting and Bayes linear graphical models are the main subjects of this section because these methods are more complex. This sections ends with a brief notion on analogous data sets and the three-point approximation.

4.1 Parametric fitting

Complex elicitation methods usually impose a structure on an expert's opinion. Often it is assumed that the knowledge of the expert can be well-represented by some member of a specified family of distributions. Members of the family are distinguished by parameters (called *hyperparameters*) and the elicitation task then reduces to choosing appropriate hyperparameter values to capture the main features of the opinions. The family of distributions is typically chosen so that it facilitates following analysis if data becomes available. For many models this does not stand in the way of the research. Usually the family offers enough flexibility and can represent a variety of opinions through suitable choice of hyperparameters.

Through elicitation the values hyperparameters can be obtained. Predefined structures link the hyperparameters to the parameters of family of distributions. In order to deal with potential 'elicitation errors', the hyperparameters are exposed to operations like averaging.

An example of an imposed structure is that of the normal linear model. The joint quantiles are first assessed and after that, they are fitted in a normal distribution with the following the prior structure:

$$Y|X, \beta, \sigma^2 \sim N(X^T\beta, \sigma^2),$$

$$\beta|\sigma^2 \sim N(b, \sigma^2 R^{-1}),$$

$$1/\sigma^2 \sim \chi^2_\delta/w\delta$$

Here $P(Y|\ X, \beta, \sigma^2)$ represents the probability of interest, Y, given the observations of X, β, σ^2. The hyperparameters b, R, w and δ are the parameters that have to be elicited. By eliciting probability judgements about quantities from the predictive distribution, the values can be derived. After this step, averaging is used to handle errors that could occur during elicitation. For a more elaborate discussion see [2] and [3].

4.2 Bayesian graphical model

A Bayesian graphical model is a representation of the joint probability distribution for a set of variables. The structure of a Bayesian graphical model is a graph consisting of nodes and edges, with the nodes representing the variables. The values of the variables are referred to as states. The edge between node A and B indicates a direct influence between the state of A and the state of B. The belief about a variable A with states $a_1, ..., a_n$ is a probability function. The specification of the Bayesian graphical model consists of a set Φ of possible functions over the sets of variables.

Bayesian network. A Bayesian network is a directed acyclic graph. A directed edge represents a causal impact. The essential property of the model is that it reflects conditional independence relations.

Definition 1 *Two states A and B are independent if knowledge of A does not change the belief about B. A and B are conditionally independent if given C they are independent whenever the state of C is known. In other words:*

$$P(A|B, C) = P(A|C).$$

In the Bayesian graphical model three types of relations are identified. These connections in the network are characterized as serial, diverging and converging, see Figure 1.

The conditional independence relations can be read from the directed acyclic graph through the following rule:

Two nodes A and B in a directed acyclic graph are independent if for all paths between A and B there is an intermediate node C such that either

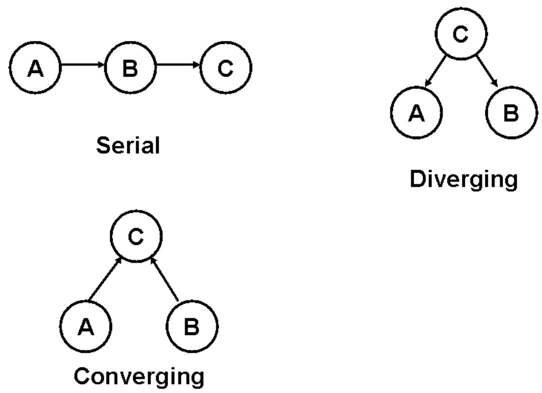

Figure 1: Relations in a graphical model.

- the connection is serial or diverging and the state of C is known
- the connection is converging and neither C nor any of its descendants have received evidence. Receiving evidence means that the information flowing through the graph reaches the node. Because the node does not necessarily receive all the information, the graph represents causal impact, the term evidence is used.

The set of potential functions Φ for a Bayesian network consists of conditional probability tables, for each variable A with parents $pa(A)$ the conditional probability $P(A|pa(A))$ is specified. The main theorem behind the use of Bayesian networks as models for reasoning under uncertainty is the following:

The chain rule for Bayesian networks. When collection U contains one or more elements of the collection of states, $P(U)$ is the product of the specified conditional probability tables: $P(U)= \Pi_{A \epsilon U}\ P(A|pa(A))$.

A direct representation of $P(U)$ may not be possible, while the size of the Bayesian network is manageably small.

For the calculations of $P(U)$ all the $P(A|pa(A)$ have to be specified. For example, in the figure of the relations, for the serial relation $P(A)$, $P(B)$, $P(C)$, $P(B|A)$ and $P(C|B)$ have to be specified. For the diverging relation $P(A)$, $P(B)$, $P(C)$, $P(B|C)$ and $P(A|C)$ have to be specified, and for the converging relation $P(A)$, $P(B)$, $P(C)$, and $P(C|A, B)$ have to be specified.

4.3 Approximations

4.3.1 Analogous data set

This method is a followup method to the ones described in the previous sections. When an analogous data set to the data set at hand exists several adjustments can be made to this existing set. These adjustments can consist of shifting the expectation or resizing the variation. Because there already is a distribution, little research has to be done about the shape or family of the distribution.

4.3.2 Three-point approximation

Another simple tactic to approach the extraction of moments is the three-point approximation. Although this method only gives estimates for the expectation and variance, it can still be very useful. This method is based on empirical observations in estimating parameters.

The approximations are all based on three of the distributions. The most commonly used points are the 0.025, 0.975, 0.05, 0.095 quantiles, the median and mode. Several relations of the mean and variance to these points are proposed.

5 Combining expert judgement in the BLM

In this section the handling of judgements of multiple experts when their opinions differ is discussed, based on [9]. In many Bayesian models expert assessments are treated as observational data, so that it requires a prior specification and a likelihood function constructed by the analyst. In problems where BLM is appropriate it is not unlikely that the analyst has little knowledge about the subject and is unable to specify a prior distribution. Also the analyst is burdened with the choice between a number of experts that differ in opinion.

When the experts have all given coherent beliefs concerning the random quantities a linear pool can be constructed with these beliefs. Using the properties of Marginalisation, Zero Preservation and Strong Setwise Function, see [4], it can be shown that specifying the linear pool is equivalent to the Strong Setwise Function Property. From this proof it is derived that if experts each provide a coherent set of expectations, the decision maker is also coherent.

5.1 Performance-based weighting

Performance-based weighting (PBW) consists of deriving weights from performance on seed variables. The performance on these seed variables is translated to a combined assessment for other variables. The seed variables should of course closely resemble the other variables. The classical model for constructing a linear pool is the model of Cooke, see [4]. This model combines PBW with probability assessment. Experts are elicited a number of quantiles about multiple random quantities, some of them have realisations known by the analyst, the seed variables. These seed variables are used to define two quantities for each of the experts, calibration score and information score. These scores are used to create a weight per expert with the property that it is a scoring rule.

Calibration score is a way of measuring the degree to which an expert is able to assess a quantile. Information score is a way of measuring the degree to which a expert can give a sharp uncertainty assessment. This information is a relative quantity depending on the position of the expert in the judgements of all experts. The product of the calibration and the information score gives a weight for each expert. All opinions of the experts whose weight falls within a predetermined interval are normalised, weights that fall outside of this interval are set equal to zero.

5.2 Weighting scheme

A base for a scoring rule can be seen as the following:

When for seed variables $X_1, ..., X_n$ the expectation is assessed by $a_1, ..., a_n$, a penalty function can be defined as

$$\phi(a_1, ..., a_n) = \sum_j c_j(x_j - a_j)^2,$$

with x_j being a realisation of random quantity X_j and c_j being the weight of the opinion of the expert. Note that $\varphi()$ can be seen as a loss model.

Now for each expert separately the value of φ is calculated. Next, a cut-off value, $\alpha \geq 0$, is chosen and each expert j with $\varphi_j \geq \alpha$ will be given the value zero for his weights. For each other expert i that remains, the score is the difference between the cut-off and the penalty, $\alpha - \varphi_i$. These values are then normalized and this results in an individual weight for each expert. The cut-off value α is a value chosen from the interval $(\min(\varphi_1, ...\varphi_n),\infty)$ so that the combined distance of all φ is minimized.

When the analyst constructs a linear pool like, this several statements can be made about the weighted scheme.

1. The expert with the smallest loss is always present in the linear pool
2. The analyst loss i.e., the scaled quadratic deviation from the realisation, is always smaller than or equal to the loss of any expert
3. The loss of the analyst is always equal to or less than the loss obtained when equal weights are applied
4. The weighted scheme provides a continuous mapping from the expert result to a vector of expert weights

6 Summary

In this paper we tried to give a comprehensive overview of the aspects involved when constructing a decision model using Bayes Linear Models (BLM). This approach can simplify the phase of extracting information of experts, in comparison to other decision support models. The main feature of BLM is the use of expectations as the key to the underlying stochastic processes. There is no need to filter out the full probability distribution of the underlying process. Using only the expectation, variance and covariance, a total model can be constructed.

The process of obtaining the information needed is, although simplified, still complicated. When this difficult process is analyzed, the only method of getting the input parameters is eliciting experts on the field at hand. The use of knowledge of experts is not new but a structured approach to obtaining the information still does not exist. When eliciting information, the elicitator has to take in account several psychological aspects, like the law of small numbers, hindsight bias and overconfidence, and neurological processes like judgement by representativeness, judgement by availability and judgement of anchoring and adjustment. In attempting to remove these influences on the elicitation process, a guideline is described by which an elicitation process should be conducted. Although this guideline can be considered standard for other (statistical) research method, for elicitation it is not.

After a structured method for eliciting information is established, several tools can be used for getting the moments of the probability distribution. In section four, three methods are discussed, ranging from precise distribution elicitation, parametric fitting, to a method focussed on the value of the expectation, the Bayesian graphical model.

When experts differ in opinion and little is known about quality of these opinions, a weighted scheme can help to produce a coherent representation of the experts opinions. By testing the assessments of the experts using a seed variable, their quality can be valued. Then a framework can be constructed in which the quality of each expert can be taken into account.

Concluding this paper, it can be said that the use of BLM can simplify the process of statistical analysis: the computations needed are less extensive than in the normal Bayesian approach. The difficulty lies in eliciting the moments. The ways of extracting these moments are not different to distribution extraction, so in many cases, especially if using parametric fitting, still an underlying probability function is elicited. When there is a method that can extract only expectation, variance and covariance, all benefits of BLM can be attained.

7 References

1. Goldstein, M. (1999). Bayes linear analysis. Encyclopaedia of Statistical Sciences, update volume 3. S. Kotz, Wiley: 29-34.

2. Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. J. Am. Statist. Ass., 75, 845-854.

3. Wolfsen, L. J. (1995). Elicitation of priors and utilities for Bayesian analysis. PhD Thesis. Department of Statistics, Carnegie Mellon University, Pittsburgh.

4. Cooke, R. M. (1991). Experts in Uncertainty: Opinion and Subjective Probability in Science. New York: Oxford University Press.

5. Cooke, R. M. and Goossens L. H. J. (2000). Procedures guide for structured expert judgement. European Commission. Report EUR 18820. Brussels-Luxembourg.

6. Finetti, B. de. (1974). Theory of probability, vol. 1. Wiley, New York.

7. Kadane, J. B. and Wolfson, L. J. (1998). Experience in elicitation, The Statistician, 47, part 1, 2-19. Carnegie Mellon University, Pittsburgh.

8. Garhwaite, P. H. and O 'Hagan, A. (2000). Quantifying expert opinion in the UK water industry: an experimental study, The Statistician, 49, part 4, 455-477.

9. Wisse, B., Bedford, T. and Quigley, J. (2005). Combining Expert Judgements in Bayes Linear Methodology. Dept. of Management Science, Strathclyde University, Glasgow, Schotland.