Rev.R.Acad.Cienc.Exact.Fis.Nat. (Esp)
Vol. 93, N.º 3, pp 329-336, 1999
Monográfico: Métodos Bayesianos en las Ciencias

BAYES LINEAR NETWORKS AND NONLINEARITIES

(Bayes linear/graphical model/nonlinearity/partial belief specification/inventory, dynamic linear model)

MALCOLM FARROW

School of Computing and Information Systems, University of Sunderland, Sunderland SR6 0DD, United Kingdom. CSOmfa@ISIS.sund.ac.uk.

ABSTRACT

Bayes linear methods are based on second order belief specifications and Bayes linear rules for belief adjustment. That is we elicit means, variances and covariances, rather than a complete probability specification, and adjust our expectations directly when data are observed. This approach, combined with graphical modelling, has been successfully applied to complex belief structures, for example in a user-friendly decision support system for management. In some applications it is convenient to express expectations about some unknowns on one scale and about others on another with a nonlinear transformation between the scales. For example, in a stock control system we might think in terms of the logarithms of sales for the sales-forecasting part of our belief specification. However, to handle the relationships between past and present stocks etc., we need to work directly in terms of the quantities themselves. A nonlinear link is thus necessary somewhere between sales and stocks. This paper describes work on an approach to constructing belief structures containing such nonlinearities and updating beliefs in them, while keeping, as far as possible, to the principle of only specifying a limited number of moments. The approach depends upon forms of weak conditional independence typically found in applications. Bayes-linear belief adjustment is used within subsets of the unknowns but stronger specifications are required at the boundaries. The extent to which known results in conditional independence and in Bayes-linear methods can be extended to this situation is discussed.

RESUMEN

Redes Bayesianas lineales y transformaciones no-lineales

Los métodos Bayesianos lineales se basan en especificación de primeros momentos y en el uso reglas Bayes lineales para su ajuste, de forma que se especifican medias, varianzas y covarianzas, en lugar de una distribución compelta, y se actualizan directamente los valores esperados de interés una vez observados los datos. Esta metodología, combinada con el uso de modelos gráficos, ha sido aplicada con éxito en la especificación de estructuras de información complejas como, por ejemplo, en la construcción de sistemas de soporte para decisiones empresariales. En algunas aplicaciones es conveniente expresar unos valores esperados en una escala y otros en otra escala que es una transformación no lineal de la primera. En este trabajo describe una forma de construir estructuras de información que contengan tal tipo de transformaciones no lineales y permita actualizar valores esperados de interés una vez observados los datos manteniendo, dentro de lo posible, el principio de especificar únicamente un número limitado de momentos. El método sólo requiere unas condiciones débiles de independencia condicional que son generalmente satisfechas en las aplicaciones.

1. INTRODUCTION

1.1. Bayes linear methods

Provided we can actually develop and analyse such specification, a standard Bayesian approach, involving a coherent belief specification in the form of a joint probability distribution over all of the unknowns, enables us to find the answers to questions of interest involving, for example, revised beliefs given information on the values of some of the unknowns. However the specification of a complete joint probability distribution might be an unreasonable demand. Even if we first identify a conditional independence structure, perhaps using an influence diagram, the conditional distributions which are still required will be extremely difficult to specify realistically. Secondly the computational problems involved in using such a belief specification, for example in evaluating high-dimensional conditional distributions, may be severe.

In the Bayes-linear approach, only prior means, variances and covariances, rather than complete probability distributions, need to be specified. These expectations relate directly to the quantities of interest and their associated uncertainties and are therefore more likely to represent genuinely held beliefs. The values are specified directly. We elicit subjective expectations of unknown quantities, variances which quantify the uncertainties in the values of the unknowns and covariances which describe the associations in the sense that learning the value of one quantity would cause a revision in the expectation of another.

When the values of some quantities become known, the expectations of the others can be adjusted according to the linear rule which minimises expected quadratic loss. The calculation of adjusted beliefs is computationally undemanding and is analogous to finding conditional expectations. However, we are not necessarily restricted to working with linear functions of the quantities of interest. Provided we are willing to specify the required mean, variance and covariances, we can include in the prior specification any function of any unknown as another unknown.

A recent case study in the application of Bayes linear methods is given in Farrow, Goldstein and Spiropoulos (1997).

1.2. Influence Diagrams

Even with only the first two moments required, making a genuine and coherent belief specification for a set of unknowns with complex interrelationships may not be easy. The task can by made easier by graphical modelling. Smith (1989) has shown that influence diagrams can be used to represent general relationships such as Bayes-linear structures where the usual conditional independence of probabilistic structures is replaced by weak conditional independence. Thus our influence diagrams differ from the more usual type. Each node represents a vector of quantities. Consider Figure 1 and suppose that the arc is directed from L_X to L_Z. In the usual type of influence diagram, the quantities represented by nodes X and L_Z would be conditionally independent given the value at L_X. The Bayes-linear interpretation of the diagram is that, in our adjusted beliefs, after linear fitting on the value of the vector at L_X, the correlations between the quantities at X and those at L_Z are zero. Also,

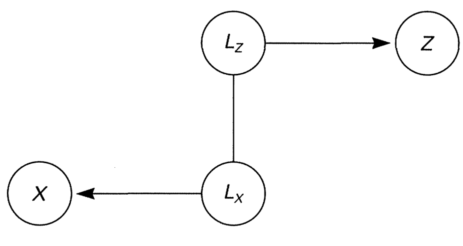

Figure 1. Basic structure.

once we know L_Z we would regard X and L_X as irrelevant to the linear prediction of Z. In what follows in this paper it will be necessary to modify both the notion of weak conditional independence and the interpretation of the diagrams.

1.3. Nonlinearities

The case study in Farrow, Goldstein and Spiropoulos (1997) concerns the production, stocks, distribution and sales of several brands of beer from a brewery. It involves a large and complicated Bayes-linear influence diagram. On one side of the diagram are production, stocks, etc. On the other side is a time-series model of beer sales. The two sides are linked since the brewery supplied the beer for the sales and the sales created the demand for the production. On the sales-forecasting side it might have been preferred to work in terms of the logarithms of sales. However the simple additive relationships on the production and stock side dictate working directly in terms of volume of beer. Thus we appear to need a nonlinear link between two collections of quantities, within each one of which we are content to use a Bayes-linear belief structure.

Consideration of the wide variety of models used in Bayesian analysis today shows that there are many situations where a nonlinear link of this type may be required.

In this paper I consider the possibility of an updating method, in the spirit of the Bayes-linear method, which can deal with such cases. Apart from Section 2, the methods considered will not actually be Bayes linear, since nonlinearities are explicitly involved. However the aim is to continue to require only a limited number of moments, generally only first and second, to be specified. Departure from Bayes linear updating means that the interpretations of weak conditional independence and of the influence diagrams will change. However, with this proviso, it remains the case that, in Figure 1, for example, once we know L_Z we would regard X and L_X as irrelevant to the prediction, by our chosen means, of Z.

In the brewery example, and in many other applications, the following conditional independence property may be exploited and, in the rest of this paper, I assume that this property holds. We can partition the unknowns into three subsets. The stocks, etc., in volume of beer, form one. The sales model, in log. volume, forms another. These are separated by a set of quantities which will be represented in the diagrams by *link nodes* which are considered with respect to both scales.

In general there is a structure as shown in Figure 1. The unknowns on the two scales are represented by $X = (x_1, ..., x_p)'$ and $Z = (z_1, ..., z_r)'$ respectively. The undireted arc between $L_X = (l_{X1}, ..., l_{Xq})'$ and $L_Z = (l_{Z1}, ..., l_{Zq})'$ represents a deterministic relationship with $l_{Zi} = g(l_{Xi})$, for $i = 1, ..., q$, for some specified, strictly monotonic, function

g. The arc is undirected since it is assumed that the conditional independences implied by a directed arc in either direction hold. Note that, with this condition, if we started with a diagram where either one (but not both) of the outer arcs was directed towards L, rather than away from L, it could be reversed to obtain Figure 1. That is the following weak conditional independences hold.

$$X \perp\!\!\!\perp L_Z, Z \mid L_X, \qquad Z \perp\!\!\!\perp L_X, X \mid L_Z.$$

2. FULLY BAYES-LINEAR APPROACH

The Bayes-linear method depends only on our ability to specify means, variances and covariances for the unknowns. Therefore, provided we are prepared to specifify the necessary moments we can include X, L_X, L_Z and Z in one Bayes-linear structure. However this will not be satisfactory to a user who wishes to make explicit use of his or her belief in a nonlinear relationship. In particular, when some data have been observed, the user's actual variances and covariances for the remaining unknowns are likely to be different from those predicted by Bayes-linear updating.

Nevertheless, consideration of how information is propagated in the Bayes-linear case may give insight into how the nonlinear case might be handled. If we interpret Figure 1 in the Bayes-linear sense then we can, for example, represent Z as

$$Z = M_Z + K_{LZ}(L_Z - M_{LZ}) + U_Z \tag{1}$$

where M_Z and M_{LZ} are the prior expectation vectors for Z and L_Z respectively, K_{LZ} is a $r \times q$ matrix and U_Z is a vector of zero-mean quantities, uncorrelated with L_Z, L_X or X. The variance matrix, V_Z, of U_Z is called a *specific variance* (see Farrow, Goldstein and Spiropoulos, 1997). Thus the variance of Z can be separated into that part associated with L_Z, i.e. $K_{LZ}\text{var}(L_Z)K'_{LZ}$, and the unexplained variance V_Z. Further, if we observe X then the mean and variance of Z change in the way indicated simply by adjusting the mean and variance of L_Z by X and applying (1). In fact, if we partition each of X, L_X, L_Z, Z into two parts, one of which, X_1, L_{X1}, L_{Z1}, Z_1, is observed, we can adjust our beliefs, in the Bayes-linear context, about the remainder, X_2, L_{X2}, L_{Z2}, Z_2, by adjusting by each of X_1, L_{X1}, L_{Z1} in any order and, for example, if we start with X_1, then we can adjust X_2, L_{X1}, L_{X2} by X_1 and calculate, stepwise, using relations similar to (1), how this propagates through to L_Z and Z.

3. GENERALISING THE BAYES-LINEAR IDEA

3.1. Features desirable in an inference system

In this section I outline the features which it would seem desirable to have if we are to generalise the Bayes-

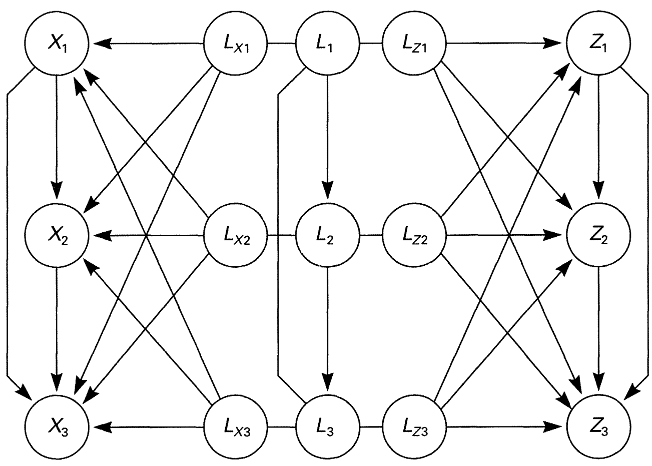

Figure 2. Unknowns in three parts.

linear approach to cope with non-linear links. I will refer to the quantities in Figure 2 in which each of X, L_X, L_Z, Z is partitioned so that, for example, $X' = (X'_1, X'_2, X'_3)$. While X_i and Z_i need not contain the same number of elements, clearly L_{Xi} and L_{Zi} must contain corresponding elements. This partition allows us to consider the effects of stepwise observation of groups of unknowns and the related conditional independence properties. Consideration of the effect of observing $Y'_i = (X'_i, L'_{Xi}, L'_{Zi}, Z'_i)$ includes cases such as observing some of X and some of the link nodes but none of Z since this would have the same effect as the case where Z_i is uncorrelated with L_Z and with the rest of Z. To clarify how the link nodes are related, I have introduced L_i between L_{Xi} and L_{Zi}. There are deterministic, possibly nonlinear, relationships between L_i and each of L_{Xi} and L_{Zi} but often, in practice, we would either have $L_{Xi} \equiv L_i$ or $L_{Zi} \equiv L_i$. Clearly, observing any one of L_{Xi}, L_i, L_{Zi} implies observing all of them and, of course, if this were a conventional probabilistic influence diagram then we would not need to show L_X or L_Z.

The underlying principle is that, with the exception of the link nodes, we only require specification of first and second moments. At the link nodes we require only those specifications which are necessary to relate the first and second moments on the two sides. In a full probabilistic specification, all moments are given and these higher order relationships can be exploited in updating. Since these higher moments are not specified here, knowledge of them is not used in updating.

R-1. If we observe all of the link nodes, L, then beliefs about unknown elements of X and Z should be updated by Bayes-linear updating by L_X and L_Z respectively. Subsequent observation of elements of X should result in Bayes-linear adjustment of beliefs about the remaining unknown elements of X (similarly Z). Once L is known, X is uninformative about Z and Z is uninformative about X.

Specifically, we can represent our beliefs about Z given L by writing

$$Z_1 = M_{Z1} + \sum_{i=1}^{3} K_{LiZ1} (L_{Zi} - M_{LZi}) + U_{Z1},$$

$$Z_2 = M_{Z2} + \sum_{i=1}^{3} K_{LiZ2}(L_{Zi} - M_{LZi}) + K_{Z1Z2}(Z_1 - M_{Z1}) + U_{Z2}, \quad (2)$$

$$Z_3 = M_{Z3} + \sum_{i=1}^{3} K_{LiZ3}(L_{Zi} - M_{LZi}) + \sum_{i=1}^{2} \mathrm{K}_{ZiZ3}(Z_i - M_{Zi}) + U_{Z3} \quad (3)$$

were the K's are appropriate matrices and U_{Z1}, U_{Z2}, U_{Z3} are uncorrelated with each other.

R-2. We need to declare some beliefs about the relationship between L_X and L_Z. Ideally this will only involve the first and second moments of each. In this case we need a 1-1 maping between (M_{LX}, V_{LX}) and (M_{LZ}, V_{LZ}) where M_{LX}, M_{LZ} are the mean vectors and V_{LX}, V_{LZ} are the variance matrices for L_X, L_Z respectively. To gain some guidance as to the sort of relationship which we might use, we might consider some probabilistic examples. However we are not supposing that such full probabilistic relationships are actually specified. We only specify the relationships among the first two moments and do not exploit other relationships which would be available if we did give a full probabilistic specification.

R-3. If information is only ever propagated in one direction, for example if we can observe some or all of X but never observe any of Z, then updating on the unobserved side takes place through the link nodes and the relationship in R-2. From (1) we see that $E(Z|X_i) = M_Z + K_{LZ}[E(L_Z|X_i) - M_{LZ}]$ and $\mathrm{var}(Z|X_i) = K_{LZ}\mathrm{var}(L_Z|X_i)K'_{LZ} + V_Z$. By R-2 we can obtain $E(L_Z|X_i)$ and $\mathrm{var}(L_Z|X_i)$ from $E(L_X|X_i)$ and $\mathrm{var}(L_X|X_i)$ and by R-1 it would appear that these latter quantities are obtained by Bayes-linear adjustment of L_X by X_i, but see Section 3.2 below.

R-4. We should be able to use a suitably defined form of generalised conditional independence satisfying the three properties below established by Dawid (1979).

CI-1 For all subsets Y_1, Y_2, Y_3, we have $Y_3 \perp\!\!\!\perp Y_2 | Y_1 \cup Y_2$.

CI-2 For all subsets Y_1, Y_2, Y_3, we have $Y_2 \perp\!\!\!\perp Y_3 | Y_1 \Leftrightarrow Y_3 \perp\!\!\!\perp Y_2 | Y_1$.

CI-3 For all subsets Y_1, Y_2, Y_3, we have

$$Y_4 \perp\!\!\!\perp Y_3 \cup Y_2 | Y_1 \Leftrightarrow \begin{cases} Y_4 \perp\!\!\!\perp Y_3 | Y_2 \cup Y_1 \\ Y_4 \perp\!\!\!\perp Y_2 | Y_1 \end{cases}$$

For the moment we interpret $X \perp\!\!\!\perp Z | Y$ as «X is uninformative about Z, within our system of inference, given knowledge of Y».

R-5. Related to R-4 is the issue of stepwise updating. That is, updating our beliefs about Y_3 by both Y_1 and Y_2 should result in the same adjusted beliefs about Y_3 as we would obtain by updating, stepwise, Y_2 and Y_3 by Y_1 then Y_3 by Y_2. Of course this is satisfied by any system of inference in which, when new information is received, we incorporate its effect by returning to the original state and adjusting by all of the information now available. However, in itself, this is not entirely satisfactory. We would like to be able to express our intermediate state in the same form as the initial and final states and not have to return to the original state for updating.

R-6. In the Bayes-linear method, adjusted expectations are calculated according to the linear rule which minimises expected quadratic loss. Doing this requires specification of just the first and second moments. By analogy, what is required here is a rule which minimises expected quadratic loss within some class of rules and which requires specification of just the moments and relationships described above, that is the first and second moments of X given L_X and Z given L_Z, the relationship in R-2 and an appropriate relationship among the elements of L_X and of L_Z.

As an illustrative example of the kind of relationship which might be used at R-2, suppose that $g(x) = e^x$. We can construct the first two moments of l_{Xi}, l_{Xj} by writing $l_{Xi} = u_{xii} + u_{xij}$ and $l_{Xj} = u_{xjj} + u_{xij}$ where u_{xii}, u_{xjj}, u_{xij} are independent. (It is sufficient that, within each of the sets $\{u_{xii}, u_{xij}, u_{xjj}\}$, $\{\exp(u_{xii}), \exp(u_{xij}), \exp(u_{xjj}), \exp(2u_{xij})\}$, $\{\exp(2u_{xii}), \exp(2u_{xij}), \exp(2u_{xjj})\}$, the elements are uncorrelated.) Write λ_{nii}, λ_{njj}, λ_{nij} for $E[u^n_{xii}]$, $E[u^n_{xjj}]$, $E[u^n_{xij}]$ and μ_{nii}, μ_{njj}, μ_{nij} for $E[\exp(nu_{xii})]$, $E[\exp(nu_{xjj})]$, $E[\exp(nu_{xij})]$. Let m_{xi}, m_{zi} be the means and v_{xii}, v_{zii} be the variances of l_{Xi}, l_{Zi} respectively and v_{xij}, v_{zij} be the covariances of l_{Xi}, l_{Xj} and l_{Zi}, l_{Zj} respectively. Then

$$m_{xi} = \lambda_{1ii} + \lambda_{1ij},$$
$$v_{xii} = \lambda_{2ii} - \lambda^2_{1ii} + \lambda_{2ij} - \lambda^2_{1ij},$$
$$v_{xij} = \lambda_{2ij} - \lambda^2_{1ij},$$
$$m_{zi} = \mu_{1ii}\mu_{1ij},$$
$$v_{zii} = \mu_{2ii}\mu_{2ij} - \mu^2_{1ii}\mu^2_{1ij},$$
$$v_{zij} = \mu_{1ii}\mu_{1jj}(\mu_{2ij} - \mu^2_{1ij}).$$

The specification of the link is then completed by giving a relationship between the λ and μ moments. Ideally this would be elicited directly but, in practice, we might be guided by the relationships in well known two-parameter distibutions. For example we might use an analogy with the normal distribution, in which case we obtain the following.

$$m_{zi} = \exp[m_{xi} + v_{xii}/2],$$
$$m_{xi} = \ln(m_{zi}) - \ln[1 + v_{zii}/m^2_{zi}]/2,$$
$$v_{zii} = \exp[2\mathrm{m}_{xi} + v_{xii}](\exp[v_{xii}] - 1),$$

$$v_{xii} = \ln[1 + v_{zii}/m_{zi}^2],$$
$$v_{zij} = \exp[m_{xi} + m_{xj} + (v_{xii} + v_{xjj})/2](\exp[v_{xij}] - 1),$$
$$v_{xij} = \ln[1 + v_{zij}/(m_{zi}m_{zj})].$$

Note, however, that even if we use an analogy with a probability distribution we do not use other features of such distributions. In particular, our belief specification is not sufficiently detailed to allow us to update the relationship as a result of observing data. Such a change can occur in a full probabilistic analysis because the form of the distribution might be altered.

3.2. The problem of updating the link nodes

Suppose we have observed $Y'_1 = (X'_1, L'_{X1}, L'_{Z1}, Z'_1)$. From (2) we see that

$$E(Z_2 \mid Y_1) = M_{Z2} + K_{L1Z2}(L_{Z1} - M_{LZ1}) + K_{Z1Z2}(Z_1 - M_{Z1}) + \sum_{i=2}^{3} K_{LiZ2}[E(L_{Zi} \mid Y_1) - M_{LZi}]$$

and

$$\mathrm{var}(Z_2 \mid Y_1) = V_{Z2} + \sum_{i=2}^{3} K_{LiZ2}\,\mathrm{var}(L_{Zi} \mid Y_1)K'_{LiZ2} + K_{L2Z2}\mathrm{covar}(L_{Z2}, L_{Z3} \mid Y_1)K'_{L3Z2} + K_{L3Z2}\mathrm{covar}(L_{Z3}, L_{Z2} \mid Y_1)K'_{L2Z3}$$

Similarly, using (3), we can find the adjusted mean and variance of Z_3 given Y_1 and the adjusted covariance of Z_2, Z_3 given Y_1, all in terms of the adjusted means, variances and covariances of L_{Z2}, L_{Z3} given Y_1. Thus the key to further progress lies in determining how to adjust beliefs about L_{Z2}, L_{Z3}.

Consider only $Y'_1 = (X_1, L_{X1}, L_1, L_{Z1}, Z_1)$ and L_{X2}, L_2, L_{Z2} in Figure 2. We need to consider what happens to the first two moments of L_{X2}, L_2, L_{Z2} when we observe Y'_1.

In standard Bayes-linear work, updating is analogous to the case of multivariate normality with known second order moments. An attempt to find a probabilistic analogy for guidance here is unlikely to be fruitful because of the need to find distributions for X and Z which would be conjugate across the nonlinear link and yet such that each would lead to Bayes-linear updating on its own. We could suppose, for example, that X is normal given L_X and Z is normal given L_Z and perhaps that L_X is itself normal. However, once some of Z became known, this normality would be lost.

A second difficulty concerns the relationships within L_X and L_Z. Learning the value of L_{X1} would imply also learning the value of L_{Z1} but we can not relate our beliefs about L_{X2} to L_{X1} by a linear rule and relate our beliefs about L_{Z2} to L_{Z1} by a linear rule and expect to obtain updated beliefs about L_{X2} and L_{Z2} which «match» each other. Thus, if we are to use a Bayes-linear style specification for the relationship among the link nodes then it should be through L. In practice this presents little difficulty as far as the specification of initial beliefs goes since, in applications, it will usually be more natural to have the linear relationship on one side, e.g. $L \equiv L_X$. It does, however, present us with another difficulty.

From R-3 we see how we might update beliefs about the link nodes given observation of either some of X or some of Z but not both. Two questions are raised. Firstly, what happens if we observe both some of X and some of Z? We would require, of course, that observing some of X and then some of Z would give the same result as observing the same data in the other order or simultaneously. Secondly, in view of the comments in the preceding paragraph, if $L \equiv L_X$, can we really use Bayes-linear updating of L_Z by Z data or does the updating have to be referred to L_X?

4. SPECIAL CASES

4.1. Introduction

The apparently simple approach of Section 3 has raised some awkward questions to which it appears we do not yet have completely satisfactory answers. In Section 5 below I will attempt to set the scene for investigation of a more general framework. Before that, in the rest of Section 4, I will describe some special cases where we might usefully apply the ideas already discussed. One of these illustrates the difficulty in trying to apply them.

4.2. Inventory prediction systems

In Section 1 I referred to the system described by Farrow, Goldstein and Spiropoulos (1997). Consider what

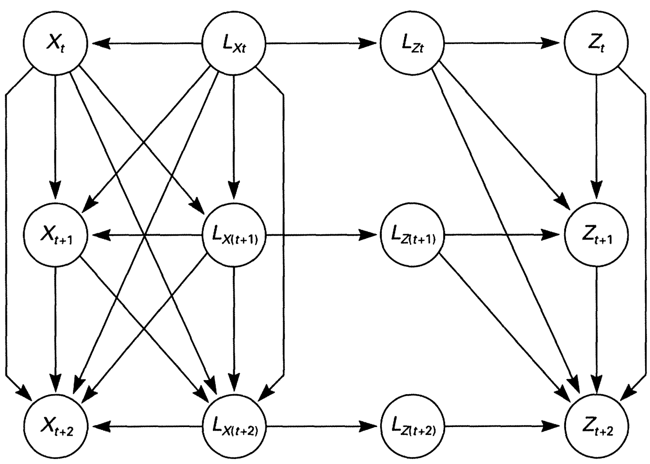

Figura 3. Inventory Forecasting.

actually happens here if we do indeed forecast the sales on a logarithmic scale. We can simplify the situation in Figure 2 to that shown in Figure 3. Here $L \equiv L_X$. We have a sequence of time steps, say weeks. At time t we know the values of all data up to and including time t. The unknowns on the left, X_t, L_{Xt}, constitute the sales-forecasting part of the system. The logarithms of the sales of the various beers in week t are in L_{Xt} while X_t might contain lagged values of L_{Xt}, trend and seasonal terms and other relevant variables. The non-logarithmic quantities, such as depot and brewery stocks, production, deliveries etc. are in Z_t which is influenced by the non-logarithmic sales L_{Zt}.

At time t our task is to forecast future values, all of which are unknown. Given X_t and L_{Xt}, all future sales values, X_{t+k}, $L_{X(t+k)}$, $L_{Z(t+k)}$, for $k > 0$, are conditionally independent of past and present Z values. We can forecast the sales, giving means, variances and covariances for $L_{X(t+1)}$, $L_{X(t+2)}$, ... These are converted, using R-2, into means, variances and covariances for $L_{Z(t+1)}$, $L_{Z(t+2)}$, ..., which are in turn fed into the forecasting of Z_{t+1}, Z_{t+2}, ... using (2), (3) etc., with K_{LiZj} only nonzero when $i \leqslant j$.

In this case we do not encounter difficulties because, in all situations which actually occur, we know how to update the link nodes.

4.3. Dynamic linear model with nonlinear link

Consider the structure of Section 4.2 with the following differences. Firstly, at time t, we know just the values of Z_{t-k} for $k = 0,1$, 2, ... That is we do not observe the unknowns on the left. Secondly, $Z_t, L_{Zt} \perp\!\!\!\perp Z_{t+j}, L_{Z(t+j)} \mid X_t$ for $j \neq 0$. That is, we delete all of the non-horizontal arcs on the right hand side of Figure 3.

This situation is rather similar to the dynamic generalised linear models of West, Harrison and Migon (1985), although we have not specified a probability distribution for Z_i and we have not restricted Z_i or, for the moment, L_{Zi} to be a scalar. Situations of this sort might arise, for example, if we have a multivariate time series where the underlying state is evolving linearly on a logarithmic scale but the observations are subject to possibly correlated errors with variances which do not depend on the means. This is also a step on the way towards being able to handle, for example, Poisson-like observations, although, for this, we would also need to be able to deal with the implied mean-variance relationship.

The idea is that we proceed as follows but see section 5.4 below. At time t we observe Z_t. This leads to Bayes-linear updating of the mean and variance of L_{Zt}. This, in turn, causes revision of the mean and variance of L_{Xt}, using R-2. This information is easily propagated through the linear structure to X_t, $L_{X(t+1)}$, X_{t+1}, $L_{X(t+2)}$, X_{t+2} etc. We can then use R-2 to calculate new means and variances for $L_{Z(t+1)}$, Z_{t+1}, $L_{Z(t+2)}$, Z_{t+2} etc. Then, at time $t+1$, we observe Z_{t+1} and so on.

The importance of the conditional independence structure and the difficulty in attempting to apply the ideas more generally are illustrated by considering updating simultaneously by Z_t and Z_{t+1}. In this case we might imagine that we could treat the data at times t and $t+1$ as a single set and update the means, variances and covariances of L_{Zt}, $L_{Z(t+1)}$ directly, by Bayes-linear adjustment. However this would lead to a different (though, perhaps, only slightly different) mean and variance for L_{Zt} compared to that obtained by stepwise updating and propagation through L_X. This difference would then be passed on through $L_{X,t+1}$. In fact this problem suggests that, if we are to be able to handle more general situations, for example observing only some of Z_t, then we will need to restrict L_{Zt} to be a scalar.

5. TOWARDS A MORE GENERAL INFERENCE METHOD

5.1. Direction of information flow

First consider straightforward Bayes-linear updating. Suppose we have just two vectors of unknowns, X, Y, with prior means M_X, M_Y, prior variances V_{XX}, V_{YY} and prior covariance $V_{XY} = V'_{YX}$. Suppose that the joint variance matrix is of full rank. Then we can always express the relationship as either $Y = M_Y + K_{XY}(X - M_X) + U_{Y|X}$ or $X = M_X + K_{YX}(Y - M_Y) + U_{X|Y}$, where $K_{XY} = V_{YX}V_{XX}^{-1}$, $K_{YX} = V_{XY}V_{YY}^{-1}$ and $U_{Y|X}$, $U_{X|Y}$ are vectors of zero-mean quantities, uncorrelated with X, Y respectively and with variances $V_{Y|X} = V_{YY} - V_{YX}V_{XX}^{-1}V_{XY}$ and $V_{X|Y} = V_{XX} - V_{XY}V_{YY}^{-1}V_{YX}$. Changing between the two representations corresponds to reversing the arc between X and Y in an influence diagram. Let us introduce the notation $X \to Y$ to denote that there is a directed arc from X to Y. Note that, if there is a third quantity Z and $Z \to X$ or $Z \to Y$ but not both, then, in order to reverse the arc between X and Y, we will need to add an extra arc so that $Z \to X$ and $Z \to Y$. (See, e.g., Smith, 1989).

Although the nonlinear links are shown as undirected arcs, the present formulation does not yet provide for bidirectional information flow across them. It is therefore necessary to treat them as directed, in effect considering moments on the «receiving» side to be conditional on the quantities on the «transmitting» side. We can then apply the normal rules for arc reversal.

5.2. Conditions for propagation

Now let us consider Figure 2 but focusing our attention on the top row, i.e. Y_1, and L_{X2}, L_2, L_{Z2}. The key issue is how we update beliefs about L_2 if we observe Y_1. Once this updating is done, propagation through the rest of the structure is straightforward. At present I know of no general method for updating L_2, although an approach slightly different to that of this paper is being investigated, but it is instructive to see how far we can get with what is known. What we can do is propagate information in the direction of the arcs. Thus we can make inferences provided that the arcs can be rearranged so that information flows in the «right» direction, without violating the basic conditions that X and Z are separated by L and that, assuming either $L_{Xi} \equiv L_i$ or $L_{Zi} \equiv L_i$, L_{Xi} is not directly connected to Z or to L_{Zj} for $i \neq j$ and L_{Zi} is not directly connected to X or to L_{Xj} for $i \neq j$.

If two of the arcs were reversed so that $X_1 \rightarrow L_{X2}$ and $Z_1 \rightarrow L_{Z2}$ then we might suppose that we could write $L_2 = M_{L2} + K_{L1L2}(L_1 - M_{L1}) + h_x[K_{X1L2}(X_1 - M_{X1})] + h_Z[K_{Z1L2}(Z_1 - M_{Z1})] + U_{L2}$, where h_X, h_Z are appropriate, possibly nonlinear, functions, with the mean and variance of $h_X(X)$ specified in terms of those of X etc. We might even suppose that the prior specification is elicited directly in this form. However this would simply transfer the problem to the propagation outwards again to X_2, Z_2.

The more natural form of prior specification is that shown in Figures 1, 2, with arcs directed outwards. Since L_{X2}, L_2, L_{Z2} are deterministically related we can, for this purpose, treat them as a single node. We can then see that we can reverse one of the arcs so that, say, $X_1 \rightarrow L_{X2}$. Unfortunately it is not then generally possible to reverse the other, so that $Z_1 \rightarrow L_{Z2}$, without adding an extra arc from X_1 to Z_1 which would violate the separation of X and Z by L. Fortunately, consideration of the special cases shows that it is sufficient in some practical problems to be able to reverse one of the arcs.

5.3. Inventory prediction systems

Here the crucial conditions are that X_{t+k}, $L_{t+k} \perp\!\!\!\perp Z_{t-j} \mid X_t, L_t$ for all $k > 0$ and $j \geqslant 0$, that Z_t is never observed before X_t and L_t, and that $L_t \equiv L_{Xt}$. Thus information flows only in the direction X to Z, with Bayes linear updating of $L_{X(t+k)}$ and updated moments of $L_{Z(t+k)}$ fed into the formulae for Z_{t+k} etc.

5.4. Dynamic linear model with nonlinear link

In this case the observations are made on the Z side and information must cross the nonlinear links in both directions. We can see whether this works as follows. First assume $L_{Xk} \rightarrow L_{Zk}$ for all k. Then do any rearrangements necessary so that no arc is directed into $L_1 \equiv L_{X1}$. This is straightforward. Having done this we can reverse the nonlinear link at time t so that $L_{Zt} \rightarrow L_{Xt}$. This amounts to thinking about our marginal beliefs about L_t in terms of L_{Zt}. Observing Z_t, or even just some of Z_t, then leads to Bayes linear updating of L_{Zt} and straightforward propagation, in the direction of the arcs, to the other unknowns, including Z_{t+1} etc.

Notice however that simultaneous reversal of two nonlinear links would cause unresolved complications. An extra arc would be required which violated the separation condition. Thus we are not able to update simultaneously by both Z_t and Z_{t+1}. It is easily confirmed that, if we attempt to do this stepwise, the results depend on the order of updating. Indeed, this argument can be applied to the links connecting the individual elements of L_{Zt} and L_{Xt} leading to the conclusion that, if we are to allow the possibility of observing some, but not all, of Z_t, L_{Zt} for a given t, we must restrict L_{Zt} to be a scalar, though Z_t need not be so restricted.

6. CONCLUSIONS

There is still much work to be done and we are still some way from a complete solution. However the examples of Section 4 show that some useful results can be obtained. In particular it seems that we can propagate information across the nonlinear link in one direction. In the inventory problem, for example, propagation is always in one direction, from sales to stocks. We are never in a situation where we would want to infer sales from observed stocks. Rather we forecast future sales from past sales and this propagates into the forecast of future stocks, etc. Similarly in the case of the dynamic linear model with a nonlinear link we obtain partial success by imposing a restriction. In this case information flows from the «nonlinear» side to the «linear» side when data are observed then flows back to give forecasts for future observations. This is done by imposing conditional independence between the data vectors, on the «nonlinear» side, given the underlying «linear» state vectors. Even with this condition, however, we are only able to update by the data at a single time step.

In Section 5 I have suggested a starting point for further investigation towards a method for belief adjustment, based on first and second moments, in more general cases. Work is under way on a more general method and I hope to be able to report on this in a future paper. At the time of writing further work is required. In particular, we need to establish that the results of any such general updating have a useful meaning and what that is. That is, the exact relationship with R-6 needs to be established.

ACKNOWLEDGEMENT

I am grateful to Michael Goldstein for a suggestion which led to a clarification.

REFERENCES

1. David, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy Statist. Soc.* **41,** 1-31.
2. Farrow, M., Goldstein, M. and Spiropoulos, T. (1997). Developing a Bayes linear decision support system for a brewery. *The Practice of Bayesian Analysis.* (S. French and J. Q. Smith eds.). Edward Arnold, 71-106.
3. Smith J. Q. (1989). Influence diagrams for statistical modelling. *Annals of Statistics* **17**, 654-672.
4. West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Amer. Statist. Ass.* **80,** 73-97.