DISTRIBUTED DIAGNOSIS IN CAUSAL MODELS WITH CONTINUOUS VARIABLES

Judea Pearl

January 1986
CSD-860051

Technical Report
R-54
CSD-8500xx
December 1985

Distributed Diagnosis in Causal Models With Continuous Variables*

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024

This work was partially supported by NSF Grant #DCR 83-13875.

1. INTRODUCTION

We consider hierarchical systems of continuous random variables, like the network depicted in Figure 1. Each variable x has a set of parent variables $u_1, u_2, \ldots, u_n$ and a set of children variables $y_1, y_2, \ldots, y_m$. The relation between x and its parents is given by the linear equation

$$x = b_1 u_1 + b_2 u_2 + \cdots + b_n u_n + w \tag{1}$$

where $b_1, b_2, \ldots, b_n$ are constants, the u's are normally distributed random variables, and w is a noise term — normally distributed, zero mean, and uncorrelated with the u's. Initially we shall assume that the network is singly connected, i.e., at most one path connects any two variables. This is equivalent to assuming that the u's are marginally uncorrelated, but become correlated once we know the value of x or any of its descendants. In Section 3 we shall relax this assumption and will treat networks with circuits.

The additive interaction of Equation (1) serves as a standard for causal models in many areas — structural equations in psychology (Bentler 1980), recursive models and path analysis in sociology and genetics (Duncan, 1975; Kenny, 1979, Li, 1975), causal models in economics (Joereskog, 1982), and others (Wright, 1921). Whereas most work in these areas focused on the issue of validating the model on the basis of empirical data, the emphasis here is on diagnosis and data interpretation. We are given the network topology, the link coefficients (b's), the variances (σ_w) of all the noise terms, and the means and variances of the root variables (nodes with no parents). Our task is to devise a distributed scheme for updating the means and variances of every variable in the network to account for evidential data D, i.e., a set of variables whose values have been determined precisely.

In principle, since the model completely specifies a correlation matrix for all variables in the system, the estimation could be performed by familiar least-square methods. However, our additional restriction is that the computation be conducted *distributedly*, as though each variable were managed by a separate, remotely located processor communicating only with processors that are adjacent to it in the network. The main reason for this computational paradigm is that it leads to a *transparent* revision process, in which the intermediate steps can be given an intuitively meaningful interpretation--a feature lacking in methods based on matrix manipulations. Since a distributed process guarantees that each computational step obtains inputs only from neighboring, semantically related variables, and since the activation of these steps proceeds along semantically familiar pathways, people find it easy to give meaningful interpretation to the individual steps and thus gain confidence in the final result. Distributed processing also makes it easier for computer systems to fortify the numerical results with qualitative justifications by tracing the sequence of operations along the activated pathways and giving them causal or diagnostic interpretations using appropriate verbal expressions.

In previous papers (Pearl, 1982; Kim and Pearl, 1983; and Pearl, 1985a) a distributed method was developed for propagating the impact of evidential data in a network of *discrete* random variables, where each variable was assumed to be related to its parents by a general conditional probability $P(x \mid u_1, u_2, \ldots, u_n)$. This paper extends the method to the case of continuous variables that interact in the manner of Eq. (1).

2. ESTIMATION IN SINGLY-CONNECTED NETWORKS

As in the treatment of discrete random variables (Pearl, 1985a), the impact of each new evidence is viewed as a perturbation that propagates through the network via message-passing between neighboring processors. Each processor x has available to it the following set of parameters:

1. the link coefficients, $b_1, b_2, \ldots, b_n$;

2. the variance σ_w of the noise which directly affects x;

3. the messages $\pi_x(u_i)$, $i=1, 2, \ldots, n$, obtained from each parent of x;

4. the messages $\lambda_{y_j}(x)$, $j=1, 2, \ldots, m$, obtained from each child of x.

$\pi_x(u_i)$ and $\lambda_{y_j}(x)$ characterize the following Gaussian conditional densities:

$$\pi_x(u_i) = f(u_i \mid D_i^+) = N(u_i; \sigma_i^+, \mu_i^+) \tag{2}$$

$$\lambda_{y_j}(x) = f(D_j^- \mid x) = N(x; \sigma_j^-, \mu_j^-) \tag{3}$$

where D_i^+ and D_j^- stand for the values of all known variables in the subnetworks connected to x via u_i and y_j, respectively. $\pi_x(u_i)$ and $\lambda_{y_j}(x)$ are reminiscent, respectively, of the prior probability and the likelihood ratio in ordinary Bayesian analysis.

Given this set of parameters, processor x must calculate the following quantities:

1. the belief distribution of variable x

$$\text{Bel}(x) = f(x \mid D) = N(x; \sigma_x, \mu_x) \tag{4}$$

where D stands for the set of all data so far observed.

2. the message $\pi_{y_j}(x)$ to be sent to y_j, the j^{th} child of x

$$\pi_{y_j}(x) = f(x \mid D - D_j^-) \quad j=1, 2, \ldots, m$$
$$= N(x; \sigma_j^+, \mu_j^+) \tag{5}$$

3. the message $\lambda_x(u_i)$ to be sent to u_i, the i^{th} parent of x

$$\lambda_x(u_i) = f(D - D_i^+ \mid u_i) \quad i=1, 2, \ldots, n$$
$$= N(u_i; \sigma_i^-, \mu_i^-) \tag{6}$$

Since all densities and messages are Gaussian, it is clear that only the means and variances need be computed and transmitted. Accordingly, our task is to compute the quantities $\sigma_x, \mu_x, \sigma_j^+, \mu_j^+, \sigma_i^-, \mu_i^-$ from the available parameters $\sigma_w, b_i, \sigma_i^+, \mu_i^+, \sigma_j^-, \mu_j^-$. This is depicted schematically in Figure 2.

Define the following parameters:

$$\sigma_\lambda = \left[\sum_j \frac{1}{\sigma_j^-}\right]^{-1} \qquad \mu_\lambda = \sigma_\lambda \sum_j \frac{\mu_j^-}{\sigma_j^-} \tag{7}$$

$$\sigma_\pi = \sigma_\omega + \sum_i b_i^2 \sigma_i^+ \qquad \mu_\pi = \sum_i b_i \mu_i^+ \tag{8}$$

Bel(x) and the messages emerging from x are given by the following three formulae (see Appendix I for derivations):

$$\sigma_x = \frac{\sigma_\pi \sigma_\lambda}{\sigma_\pi + \sigma_\pi} \qquad \mu_x = \frac{\sigma_\pi \mu_\lambda + \sigma_\lambda \mu_\pi}{\sigma_\pi + \sigma_\lambda} \tag{9}$$

$$\sigma_j^+ = \sigma_x \Big|_{\sigma_j^- \to \infty} = \left[\frac{1}{\sigma_\pi} + \sum_{k \neq j} \frac{1}{\sigma_k^-}\right]^{-1} \qquad \mu_j^+ = \mu_x \Big|_{\sigma_j^- \to \infty} = \frac{\sum_{k \neq j} \frac{\mu_k^-}{\sigma_k} + \frac{\mu_\pi}{\sigma_\pi}}{\sum_{k \neq j} \frac{1}{\sigma_k^-} + \frac{1}{\sigma_\pi}} \tag{10}$$

$$\sigma_i^- = \frac{1}{b_i^2}\left[\sigma_\lambda + \sigma_\omega + \sum_{k \neq i} b_k^2 \, \sigma_k^+\right] \qquad \mu_i^- = \frac{1}{b_i}\left[\mu_\lambda - \sum_{k \neq i} b_k \, \mu_k^+\right] \tag{11}$$

A few qualitative features of this updating scheme are worth noting. First, in the absence of any evidential data ($D=\emptyset$, $\sigma_\lambda=\infty$), the means and variances of all variables can be computed by a simple path-tracing method, using Eq. (8). For example, the mean of variable x is equal to the weighted sum of the means of all its root ancestors, and the weights are given by the products of the b coefficients along the corresponding paths, independent of the noise variances. The variance of x, likewise, is given by a weighted sum of all the noise variances along the paths connecting x to its roots. Second, the impact of an observed variable y on any of its descendants x is equivalent to cutting off the network above y and regarding y as a zero-variance root of x, with mean equal to the observed value of y. This impact is unaffected by the noise along the path from y to x. Third, the impact of an observed variable on its ancestors does depend on the noise along the connecting paths. For example, Eq. (7) describes μ_x as a linear mixture of the means of its ancestors (μ_π) and its descendants (μ_λ) with the weights determined by the corresponding variances; the lower the variance the higher the weight of influence. Finally, the minus sign in the expression of μ_i^- (Eq. (11)) captures the "explaining away" effect of interacting causes (Kim and Pearl, 1983); the more evidence we have in favor of alternative causes (high μ_k, $k \neq i$), the less we are able to attribute an observed effect (say an increase of x) to any particular cause, say μ_i.

3. COPING WITH CIRCUITS

The propagation scheme described in the preceding section will not be valid when the underlying network contains circuits, i.e., when two or more nodes possess both common descendants and common ancestors. If we ignore the existence of circuits and permit the nodes to communicate messages as if the network were singly-connected, messages might circulate indefinitely around the circuits, and the process normally will not converge to the correct equilibrium.

A propagation method that works well when the number of circuits is small is called *conditioning* and is based on our ability to change the connectivity of a network and render it singly-connected by instantiating a selected group of variables. The use of conditioning in the case of discrete variables is described in Pearl (1985b), and this Section extends the method to models containing continuous variables. We first describe the procedure involved, and leave the derivations to Appendix II.

Assume that instantiating one variable, x_k, is enough to break all circuits in the network. In other words, holding the value of x_k constant would permit us to update all variables by applying the one-pass propagation scheme of the preceding section. Instead of propagating the usual set of λ-π parameters we now require that two such sets be propagated — λ^0–π^0 and λ^1–π^1 corresponding to holding the value of x_k at 0 and 1, respectively. The computations involved under these two conditions are the same, except that in the former case each son of x_k will receive the pair of messages $\sigma_k^+ = 0$ and $\mu_k = 0$ while in the latter, the pair will be $\sigma_k^+ = 0$ and $\mu_k = 1$ (see Eq. (2)). Similarly, each parent of x_k will receive the usual σ^- and μ^- messages specified in Eq. (11), with two different settings of σ_λ and μ_λ; ($\sigma_\lambda = 0$, $\mu_\lambda = 0$) represents the case where $x_k = 0$ while ($\sigma_\lambda = 0, \mu_\lambda = 1$) represents the case where $x_k = 1$.

Let e stand for all the data so far observed. Every node x_i in the network should possess the following set of parameters:

1. $\sigma^0_{i|e} = Var\ (x_i \mid e, x_k = 0)$ (12)

2. $\mu^0_{i|e} = E\,(x_i \mid e, x_k = 0)$ (13)

3. $\mu^1_{i|e} = E\,(x_i \mid e, x_k = 1)$ (14)

4. $\mu_{k|e} = E\,(x_k \mid e)$ (15)

5. $\sigma_{k|e} = Var\ (x_k \mid e)$ (16)

The first three parameters are available from the incoming conditioned messages λ_0–π_0 and λ^1–π^1 and are the defining parameters for Bel(x_i), as in Eqs. (4) and (9) (note that $Var\,(x_i \mid e, x_k{=}1) = \sigma^0_{i|e}$). The fourth and fifth parameters, $\mu_{k|e}$ and $\sigma_{k|e}$, are assumed to be computed by the last variable to be instantiated and to be transmitted to all other nodes as attachments to the usual λ–π parameters. We shall now show how the processor at x_i computes the total belief distribution Bel$(x_i) = f\,(x_i \mid e)$ and how it should revise the value of $\mu_{k|e}$ and $\sigma_{k|e}$ in case x_i is observed and found to have the value $x_i = v_i$. These revised parameters, denoted $\sigma_{k|e'}$ and $\mu_{k|e'}$, are then appended to the set of λ-π messages emerging from x_i and thus get transmitted to all other variables.

$\text{Bel}(x_i)$ is computed from (12)-(16) by the following formulae:

$$\text{Bel}(x_i) = N\left[x_i;\, \sigma_{i\,|e},\, \mu_{i\,|e}\right] \tag{17}$$

where

$$\sigma_{i\,|e} = \sigma^0_{i\,|e} + \left[\mu^1_{i\,|e} - \mu^0_{i\,|e}\right]^2 \sigma_{k\,|e} \tag{18}$$

$$\mu_{i\,|e} = \mu^0_{i\,|e} + \left[\mu^1_{i\,|e} - \mu^0_{i\,|e}\right] \mu_{k\,|e} \tag{19}$$

When x_i is instantiated to $x_i = v_i$, its processor computes and transmits to neighboring nodes the following pair of parameters

$$\sigma_{k\,|e'} = \left[\frac{1}{\sigma_{k\,|e}} + \frac{\mu^1_{i\,|e} - \mu^0_{i\,|e}}{\sigma^0_{i\,|e}}\right]^{-1} \tag{20}$$

$$\mu_{k\,|e'} = \sigma_{k\,|e'} \left[\frac{\mu_{k\,|e}}{\sigma_{k\,|e}} + \frac{\left[v_i - \mu^0_{i\,|e}\right]\left[\mu^1_{i\,|e} - \mu^0_{i\,|e}\right]}{\sigma^0_{i\,|e}}\right] \tag{21}$$

This pair will eventually be transmitted to all other variables and will be used to compute their overall belief densities from the conditioned parameters σ^0, μ^0,μ^1, as in (18) and (19).

If the network requires the removal of K nodes ($K>1$) to break up all circuits, the updating equations become more complicated; they involve operations on $K \times K$ matrices and the transmission of K-vectors as messages. When K is large we lose the transparency provided by the message-passing scheme and global matrix manipulating techniques may be the only diagnosis tool applicable.

CONCLUSIONS

We have shown that distributed diagnosis can be conducted in causal models with continuous (Gaussian) variables, if the number of circuits is relatively small. Each step in the diagnosis process obtains inputs from semantically related variables and can be activated asynchronously. These features yield the following advantages when used in expert systems:

1. Transparency — The intermediate steps are psychologically meaningful, and can be given sound verbal explanations.

2. Flexible control — A coherent equilibrium will be reached under any activation strategy (e.g., goal driven or data driven, parallel or sequential). There is no need to keep track of which part of the network has already been updated.

REFERENCES

Bentler, P.N. (1980) "Multivariate Analysis with Latent Variables," *Review of Psychology* 31: 419-456.

Duncan, O.D. (1975) *Introduction to Structural Equation Models*. New York: Academic Press.

Joereskog, Karl G. (1982) *Systems Under Indirect Observation*. Amsterdam: North-Holland Publishing Company.

Kenny, D.A. (1979) *Correlation and Causality*. John Wiley and Sons.

Kim, J. and Pearl, J., (1983), "A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems," *Proceedings of IJCAI-83*, pp. 190-193.

Li, Ching C. (1925) *Path Analysis: A Primer*. Pacific Grove, CA: The Boxwood Press.

Pearl, J., (1982), "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach," *Proc. of AAAI Conference on Artificial Intelligence*, Pittsburgh, PA, pp. 133-136.

Pearl, J. (1985a) "Fusion, Propagation, and Structuring in Bayesian Networks," Technical Report, CSD-850016, Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles.

Pearl, J. (1985b) "A Constraint-Propagation Approach to Probabilistic Reasoning," *Proceedings of Workshop on Uncertainty and Probability in Artificial Intelligence* (UCLA, 14-16 August, 1985), pp. 185-195.

Wold, H. (1964) *Econometric Model Building*. Amsterdam: North-Holland Publishing Co.

Wright, S. (1921) "Correlation and causation," *Journal of Agricultural Research* 20: 557-585.

APPENDIX I

Eqs. (9)-(11) are derived using the following seven formulae which facilitate the manipulation of normal distributions:

$$N(x;\sigma,\mu) = \text{const. } \exp\left\{-\frac{1}{2\sigma}(x-\mu)^2\right\} \tag{I-1}$$

$$N(x;\sigma,\mu) = N(\mu;\sigma,x) \tag{I-2}$$

$$N(ax+b;\sigma,\mu) = N\left[x;\frac{\sigma}{a^2};\frac{\mu-b}{a}\right] \tag{I-3}$$

$$N(x;\sigma_1,\mu_1)\cdot N(x;\sigma_2,\mu_2) = N\left[x;\frac{\sigma_1\sigma_2}{\sigma_1+\sigma_2},\frac{\sigma_2\mu_1+\sigma_1\mu_2}{\sigma_1+\sigma_2}\right] \tag{I-4}$$

$$\prod_i N(x;\sigma_i,\mu_i) = N\left[x;\left[\sum_i\frac{1}{\sigma_i}\right]^{-1},\frac{\sum_i\frac{\mu_i}{\sigma_i}}{\sum_i\frac{1}{\sigma_i}}\right] \tag{I-5}$$

$$\int N(y;\sigma_1,\mu_1)\,N(y;\sigma_2,x)\,dy = N(x;\sigma_1+\sigma_2,\mu_1) \tag{I-6}$$

$$\int_{x_1}\cdots\int_{x_n}\prod_{i=1}^{n}N(x_i;\sigma_i,\mu_i)N\left[\sum_{i=1}^{n}b_ix_i;\sigma,\mu\right]dx_1\cdots dx_n = N\left[0;\sigma+\sum_{i=1}^{n}b_i^2x_i,\mu-\sum_{i=1}^{n}b_i\,\mu_i\right] \tag{I-7}$$

To compute Bel(x) we partition the data D into two components, D_x^+ and D_x^-, representing, respectively, data in the subnetworks above and below x.

$$\text{Bel}(x) = f(x\mid D_x^+,D_x^-) = \alpha\, f(x\mid D_x^+)\,f(D_x^-\mid x)$$

$$= \alpha\,\pi(x)\cdot\lambda(x)$$

$$\pi(x) = f(x\mid D_x^+) = \int_{u_1}\cdots\int_{u_n} f(x\mid D_x^+,u_1\cdots u_n)\,f(u_1\cdots u_n\mid D_x^+)\,du_1\cdots du_n$$

$$= \int_{u_1} \cdots \int_{u_n} f(x \mid u_1 \cdots u_n) \prod_{i=1}^{n} f(u_i \mid D_i^+)\, du_1 \cdots du_n$$

$$= \int_{u_1} \cdots \int_{u_n} N\left[x; \sigma_\omega, \sum_{i=1}^{n} b_i\, u_i\right] \prod_{i=1}^{n} N(u_i; \sigma_i^+, \mu_i^+)\, du_1 \cdots du_n$$

(and, using I-7)

$$\pi(x) = N\left[0; \sigma_\omega + \sum b_i^2\, \sigma_i^+, x - \sum_{i=1}^{n} b_i\, \mu_i^+\right] = N\left[x; \sigma_\omega + \sum_{i}^{n} b_i^2\, \sigma_i^+, \sum_{i=1}^{n} b_i\, \mu_i^+\right]$$

$$= N(x, \sigma_\pi, \mu_\pi)$$

where σ_π and μ_π are defined in Eq. (8). Similarly, for $\lambda(x)$ we have:

$$\lambda(x) = f(D_x^- \mid x) = f(D_1^-, D_2^- \cdots D_m^- \mid x) = \prod_j f(D_j^- \mid x) = \prod_j \lambda_j(x)$$

$$= N(x; \sigma_\lambda, \mu_\lambda)$$

where σ_λ and μ_λ are defined in (7). Combining these two results, and using (I-4), we obtain

$$\mathrm{Bel}(x) = N(x; \sigma_\pi, \mu_\pi)\, N(x; \sigma_\lambda, \mu_\lambda) = N\left[x; \frac{\sigma_\pi \sigma_\lambda}{\sigma_\pi + \sigma_\lambda}, \frac{\sigma_\pi \mu_\lambda + \sigma_\lambda \mu_\mu}{\sigma_\pi + \sigma_\lambda}\right]$$

which proves Eq. (9).

To find $\pi_{y_j}(x)$, we note that it is conditioned on all data except a subset D_j^- of variables which connect to x via y_j. Therefore:

$$\pi_{y_j}(x) = f(x \mid D - D_j^-) = \mathrm{Bel}(x \mid D_j^- = \varnothing)$$

$$= \mathrm{Bel}(x) \Big|_{\sigma_j \to \infty} = \mathrm{Bel}(x) \Big|_{\sigma_\lambda = \left[\sum_{k \neq j} \frac{1}{\sigma_k^-}\right]^{-1}}$$

$$= N(x; \sigma_{y_j}^+, \mu_{y_j}^+)$$

where:

$$\sigma_{y_j}^{+} = \frac{\sigma_\pi \left[\sum_{k \neq i} \frac{1}{\sigma_k^-}\right]^{-1}}{\sigma_\pi + \left[\sum_{k \neq j} \frac{1}{\sigma_k^-}\right]^{-1}} = \left[\frac{1}{\sigma_\pi} + \sum_{k \neq j} \frac{1}{\sigma_k^-}\right]^{-1}$$

$$\mu_{y_j}^{+} = \frac{\sum_{k \neq j} \frac{\mu_k^-}{\sigma_k} + \frac{\mu_\pi}{\sigma_\pi}}{\sum_{k \neq j} \frac{1}{\sigma_k^-} + \frac{1}{\sigma_\pi}}$$

That establishes Eq. (10).

To compute $\lambda_x(u_i)$ we partition the data D into its disjoint components D_i^+, $i=1, ..., n$, and D_j^-, $j=1, ..., m$, and condition $\lambda_x(u_i)$ on all parents of x. For notational convenience we temporarily denote u_i by u, b_i by b, and let the other parents be indexed by k, ranging from 1 to some n.

$$\lambda_x(u) = f(D - D_u^+ \mid u)$$

$$= \int_{u_1} \cdots \int_{u_n} \int_x f(D_1^+ \cdots D_n^+, D_1^- \cdots D_m^- \mid u_1 \cdots u_n, x, u) f(u_1 \cdots u_n, x \mid u) dx\, du_1 \cdots du_n$$

$$= \int_{u_1} \cdots \int_x \prod_j \lambda_{y_j}(x) \prod_k f(D_k^+ \mid u_k) f(x \mid u, u_1 \cdots u_n) f(u_1 \cdots u_n \mid u)\, dx\, du_1 \cdots du_n$$

$$= \int_{u_1} \cdots \int_{u_n} \int_x \lambda(x) \prod_k \frac{f(u_k \mid D_k^+) f(D_k^+)}{f(u_k)} \cdot \prod_k f(u_k) f(x \mid u, u_1 \cdots u_n)\, dx\, du_1 \cdots du_n$$

$$= C \int_{u_1} \cdots \int_{u_n} \int_x \lambda(x) \prod_{k=1}^{n} \pi_x(u_k) f(x \mid u, u_1 \cdots u_n)\, dx\, du_1 \cdots du_n$$

where

$$\lambda(x) = N(x;\, \sigma_\lambda, \mu_\lambda) \qquad \pi_x(u_k) = N(u_k;\, \sigma_k^+, \mu_k^+)$$

$$f(x \mid u, u_1 \cdots u_n) = N(x;\, \sigma_w, bu + \sum_{k=1}^{n} b_k \mu_k)$$

Using (I-2)-(I-7) we write:

$$\lambda_x(u) = \int_{u_1} \cdots \int_{u_n} \int_x N(x; \sigma_\lambda, \mu_\lambda) \prod_{k=1}^{n} N(u_k; \sigma_k^+, \mu_k^+) N(x; \sigma_\omega, bu + \sum_{k=1}^{n} b_k u_k)\, dx\, du_1 \cdots du_n$$

$$= \int_{u_1} \cdots \int_{u_n} \prod_{k=1}^{n} N(u_k; \sigma_k^+, \mu_k^+) N(bu + \sum_{k=1}^{n} b_k u_k; \sigma_\lambda + \sigma_\omega, \mu_\lambda)\, du_1 \cdots du_n$$

$$= \int_{u_1} \cdots \int_{u_n} \prod_{k=1}^{n} N(u_k, \sigma_k^+, \mu_k^+) N\left[\sum_{k=1}^{n} b_k u_k; \sigma_\lambda + \sigma_\omega, \mu_\lambda - bu\right] du_1 \cdots du_n$$

$$= N(0; \sigma_\lambda + \sigma_\omega + \sum_k b_k^2\, \sigma_k^+, \mu_\lambda - bu - \sum_k \mu_k^+ b_k)$$

$$= N\left[b\mu; \sigma_\lambda + \sigma_\omega + \sum_{k=1}^{n} b_k^2\, \sigma_k^+, \mu_\lambda - \sum_{k=1}^{n} b_k\, \mu_k^+\right]$$

$$= N\left[u; \frac{1}{b^2}\left[\sigma_\lambda + \sigma_\omega + \sum_{k=1}^{n} b_k^2\, \sigma_k^+\right], \frac{1}{b}\left[\mu_\lambda - \sum_{k=1}^{n} b_k\, \mu_k^+\right]\right]$$

Therefore, for the i^{th} parent, u_i, we have:

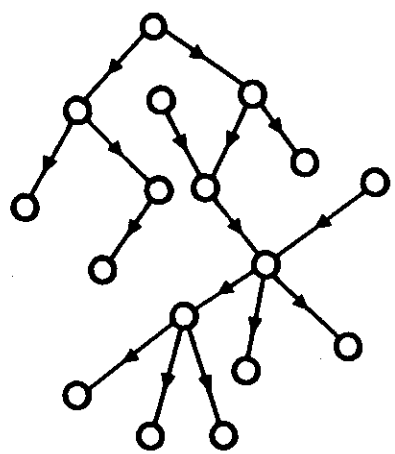
$$\lambda_x(u_i) = N\left[u_i, \frac{1}{b_i^2}\left[\sigma_\lambda + \sigma_\omega + \sum_{k \neq i} b_k^2\, \sigma_k^+\right], \frac{1}{b_i}\left[\mu_\lambda - \sum_{k \neq i} b_k\, \mu_k^+\right]\right]$$

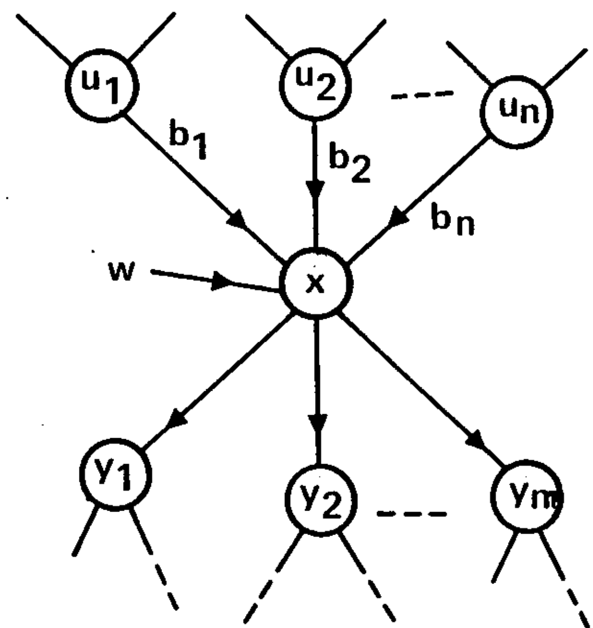
$$= N(u_i; \sigma_x^-(u_i), \mu_x^-(u_i))$$

This confirms Eq. (11) and, substituting (7), we obtain

$$\sigma_x^-(u_i) = \frac{1}{b_i^2}\left[\sigma_\omega + \sum_{k \neq i} b_k^2\, \sigma_k^+ + \left[\sum_j \frac{1}{\sigma_j^-}\right]^{-1}\right]$$

$$\mu_x^-(u_i) = \frac{1}{b_i}\left[\frac{\sum_{j=1}^{m} \frac{\mu_i^-}{\sigma_j^-}}{\sum_{j=1}^{m} \frac{1}{\sigma_j^-}} - \sum_{k \neq i} b_k \mu_k^+\right].$$


(a)



(b)

Figure 1

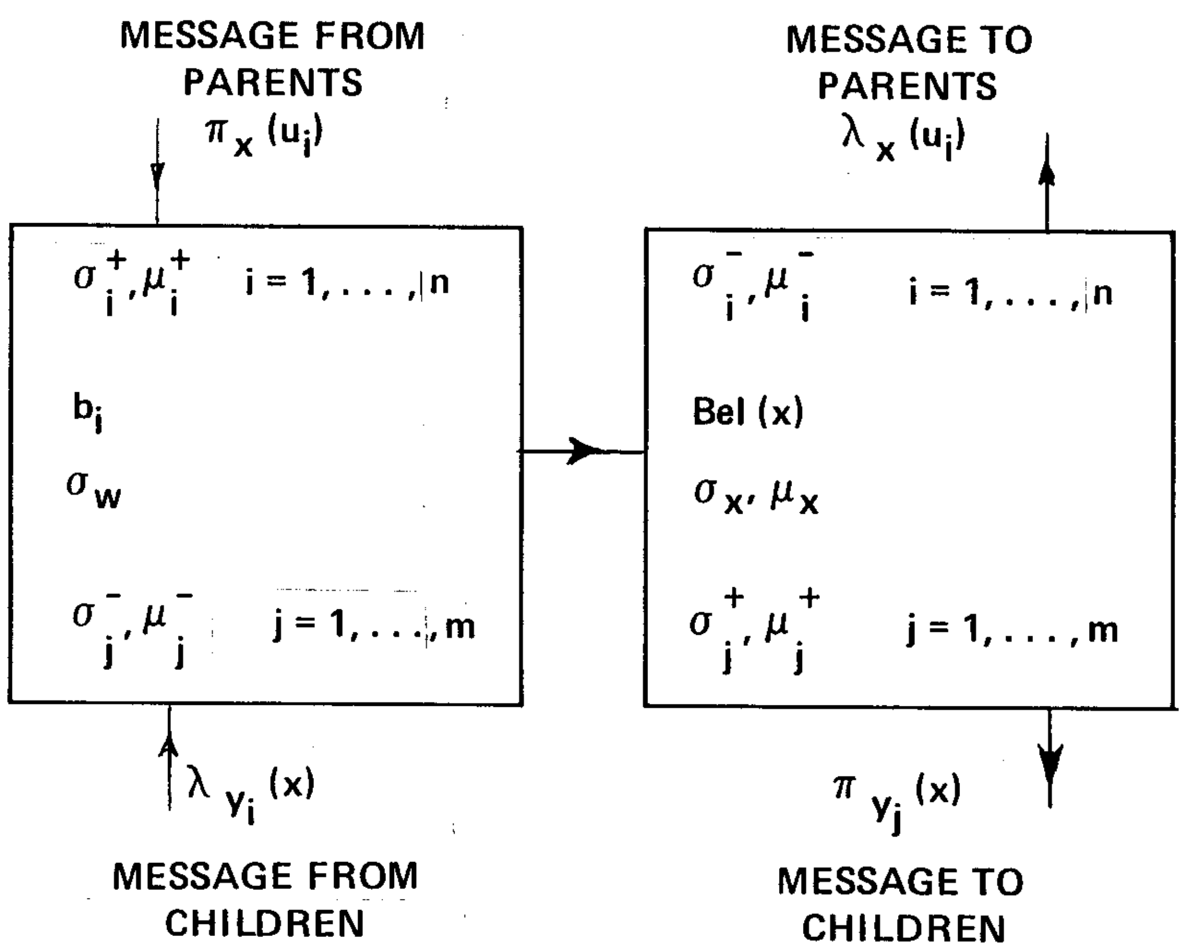

MESSAGE FROM PARENTS
$\pi_x(u_i)$
σ_i^+, μ_i^+ $i = 1, \ldots, n$
b_i
σ_w
σ_j^-, μ_j^- $j = 1, \ldots, m$
$\lambda_{y_i}(x)$
MESSAGE FROM CHILDREN
MESSAGE TO PARENTS
$\lambda_x(u_i)$
σ_i^-, μ_i^- $i = 1, \ldots, n$
Bel (x)
σ_x, μ_x
σ_j^+, μ_j^+ $j = 1, \ldots, m$
$\pi_{y_j}(x)$
MESSAGE TO CHILDREN

Figure 2