SIMULATION APPROACHES FOR CALCULATIONS IN DIRECTED GRAPHICAL MODELS

Constantin T. Yiannoutsos
Alan E. Gelfand

TECHNICAL REPORT No. 475
OCTOBER 12, 1993

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

SIMULATION APPROACHES FOR CALCULATIONS IN DIRECTED GRAPHICAL MODELS

Constantin T. Yiannoutsos
Alan E. Gelfand

TECHNICAL REPORT No. 475
OCTOBER 12, 1993

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

Simulation Approaches for Calculations in Directed Graphical Models

Constantin T. Yiannoutsos

and

Alan E. Gelfand*

ABSTRACT

In formulating models for a complex system graphical representation is an effective tool. When the components of the system are viewed as random variables, directed graphical models detail the nature of the dependence among them. Moreover, if for each variable the conditional distribution is provided according to the graph, the joint distribution is uniquely determined. Natural questions arise about the static behavior of the system under such specification as well as its response to information (observed levels of some of the variables). Answers to these questions require the ability to calculate arbitrary marginal and conditional distributions. In complex cases (high dimensional structures) such calculations require high dimensional integrations and/or summations. Most of the work to date has taken advantage of properties of directed graphs to facilitate exact calculations but is limited with regard to distributional assumptions and feasible system size. Monte Carlo methods for such calculations can accommodate much larger system size with arbitrary dependence structure and distributional forms yielding approximations which can be as accurate as desired. It is the objective of this paper to detail such methodology. An illustration is provided using a diagnostic system for congenital heart disease in neonates.

KEY WORDS

Conditional independence, Gibbs sampler, likelihood weighting, Monte Carlo.

*Constantin T. Yiannoutsos has just completed his Ph.D. in the Dep't of Statistics, University of Connecticut, Storrs, CT 06269. Alan E. Gelfand is Professor of Statistics in the same department. Gelfand's research was supported in part by NSF grant DMS 8918563.

1. Introduction.

In formulating models for a complex system graphical representation is a useful too. For non–deterministic systems, a directed graphical form provides a depiction of the joint distribution of the random components as well as clarification of the nature of the dependence structure present among them. More precisely, in such graphs, random variables are pictured by circles or dots called the *nodes* of the graph, while direct dependencies between pairs of variables are represented by arrows. The node from which an arrow emanates is the "parent", while the receiving one is the "child". Thus the graph portrays qualitative dependence amongst the variables. The induced dependence structure is quantified by specification of conditional distributions at each node given its parents. Conditional densities at nodes are denoted generically by f(X|Y) where X denotes the variable defined at the node, and Y its parents. Figure 1 shows an illustrative six node dependence structure.

[Insert figure 1 about here]

The provision of a conditional distribution for each variable in terms of its parents is sufficient to uniquely determine the distributional model for the entire graph (see e.g. Whittaker, 1990, chapter 3). We can construct the joint structure, from information provided locally without requiring profound understanding of the whole system. This facilitates construction of the model and provides a powerful communication mechanism between statisticians and their interdisciplinary audience. Most significantly however, graphical representation of the statistical model distills dependence to necessary parts. For example consider a "top down" factorization of the joint distribution in figure 1.

$$f(a,b,\ldots,f) = f(f|a,b,c,d,e)f(e|a,b,c,d)f(d|a,b,c)f(c|a,b)f(b|a)f(a). \qquad (1.1)$$

What does the graphical model tell us about the terms on the right hand side?

Recall that two (possibly vector valued) random variables X_1 and X_2 are conditionally independent given X_3, denoted by $X_1 \perp\!\!\!\perp X_2 | X_3$, iff $f(x_1,x_2|x_3) = f(x_1|x_3)\, f(x_2|x_3)$, or, equivalently, iff, $f(x_1|x_2,x_3) = f(x_1|x_3)$ and $f(x_2|x_1,x_3) = f(x_2|x_3)$. For example, in a Markov chain where dependence at state t given all preceding states $1,2,\ldots,t-1$ is limited to the immediately preceding state $t-1$,

$$f(x_t|x_{t-1},x_{t-2},\ldots,x_0) = f(x_t|x_{t-1}), \text{ i.e., } X_t \perp\!\!\!\perp (X_{t-2},\ldots,X_1)|X_{t-1}.$$

In response models for Y suppose that from the collection of explanatory variables $\underline{x}$ subset $\underline{x}_a$ is sought such that, given this subset, dependence on the rest of the variables is minimal. Then $f(y|\underline{x}) = f(y|\underline{x}_a)$, i.e. $Y \perp\!\!\!\perp X_a^c | X_a$ where X_a^c is the complement of X_a.

Returning to (1.1), the graph implies the following conditional independence relationships: $F \perp\!\!\!\perp (A,B,E)|(C,D)$ from which $f(f|a,b,c,d,e) = (f|c,d)$; $E \perp\!\!\!\perp (A,B,D)|C$ from which $f(e|a,b,c,d) = f(e|c)$; $D \perp\!\!\!\perp C|(A,B)$ which implies $f(d|a,b,c) = f(d|a,b)$; and $A \perp\!\!\!\perp B$ which implies, $f(b|a) = f(b)$. Hence (1.1) simplifies to $f(a,b,c,d,e,f) = f(f|c,d)\, f(e|c)\, f(d|a,b)\, f(c|a) \cdot f(b) \cdot f(a)$.

Effective use of the graphical model requires the ability to compute arbitrary marginal and conditional distributions. At the system level, the need to be able to calculate desired distributions arises as we try to understand quantitatively both static and dynamic system behavior. In some cases, such distributions can be used for model assessment, in the sense, of determining the plausibility of our proposed model of the system (see section 3). Once comfortable with this model, such distributions enable use of the system for diagnostic purposes. In this regard we seek to *condition* given observed information, i.e., to adjust distributions by propagation of observed information through the network. In the absence of information this amounts to marginalization. Essential here is that very little if any restriction should be imposed on data acquisition, i.e., on which part of the graphical model is considered given (fixed).

The directed graphical models we employ here have been called causal models (Pearl, 1987), graphical association models (Lauritzen, 1990a), and, most recently, causal

probabilistic networks, CPN's, (Jensen, Lauritzen and Olesen, 1991). Among the many areas of applications of these models are expert systems, artificial intelligence, genetics, reliability trees and organizational structures. Motivation for a great deal of work in the statistical uses of graphical modeling can be traced to the seminal paper of Darroch, Lauritzen and Speed (1980) on the analysis of log–linear models. This article introduced the class of graphical models, exploring in an elegant and powerful way the implications of conditional independence.

The work of Lauritzen and Spiegelhalter (1988), renewed interest in graphical approaches by developing an efficient procedure for the analysis of general discrete models. They assume that the dependence structure is completely known, i.e., the conditional probabilities of an event given all influencing factors are specified a priori, either relying on expert opinion or case specific data. Subsequent work (e.g., Spiegelhalter and Cowell, 1991) allows for imprecision in the model specification by modeling the conditional distributions at each of the discrete nodes as parametric families. The dependence structure is kept intact, while at each variable prior distributions are added to the parameters of the conditional distribution defined on it. These prior distributions are represented by parental nodes at the top of the network. Figure 2 extends figure 1 in this fashion. Extending the graphical model in this way seems sensible. The resultant joint distribution incorporates the parameters in a natural hierarchical fashion. It also enables us to maintain the assumption that the dependence structure in the graphical model is completely specified.

[Insert figure 2 about here]

An important parallel development is work by Lauritzen (1990a,b) and Wermuth and Lauritzen (1990) which formulates the Conditional Gaussian (CG) family of distributions, in an attempt to analyze mixed models i.e. models having both discrete and

continuous nodes. Though an important extension of existing theory, it is very restrictive in not allowing conditional–to–discrete dependence, i.e. no continuous node can have a discrete child

Based upon triangulation of the graph, Lauritzen and Spiegelhalter (1988) presented an algorithmic approach for analytic calculations in a purely discrete model which requires computation only at the local level. Approximate versions for mixed models are described in Lauritzen (1990b). Though elegant these methods do not offer a unified approach to the processing of general graphical models. Distributional assumptions are limited. Tailored approximations are ultimately necessary in cases where modeling makes analytical calculations intractable as is the case with non–conjugate models. High model dimension can severely strain computational resources. Monte Carlo approaches offer the promise of accommodating these problems.

Monte Carlo approaches have been considered by earlier researchers though exclusively for binary nodes. Pearl (1987) uses the neighborhood structure inherent in the network to construct the complete conditional distribution of each node. Under mild conditions the complete conditional distributions uniquely determine the joint, (e.g. Besag, 1974). Appropriate sampling of them produces a Markov chain whose stationary distribution is the one we desire. Explicit details are given in section 2.3 where we generalize Pearl's approach. Another example is the Likelihood Weighting method (Shachter and Peot, 1989). This technique uses a combination of conditional independence relations and an importance sampling scheme to derive estimates of statistics or functions of the variables from samples generated according to the discrete joint distribution. Extension to general mixed models is the subject of section 2.2.

The key point in our work is a change in focus. Rather than customary analytical attempts to obtain features of a high dimensional joint distributions we propose the use of sampling–based approaches to approximate such features. In concert with subgraph approximations, which we report on in a subsequent paper, Monte Carlo technology can

process very large systems in a unified manner, with the prospect of "black box" software (we note the current development of BUGS – Bayesian Inference Using Gibbs Sampling – as reported in Thomas, Spiegelhalter and Gilks, 1991).

In section 2 we fully describe the independent or importance sampling approach as well as the Markov chain (Gibbs Sampling) approach. In section 3 we take up classes of distributional families which are attractive as nodes in our directed graphical model, as well as the question of model assessment. In section 4, we illustrate both approaches with the analysis of a twenty–node mixed graphical model for diagnosis of congenital heart disease in neonates. We conclude this section with some notation and a few formal definitions which will be used in the remainder of the paper.

Central to our graphical methodology is the concept of a *Directed Acyclic Graph* (DAG). A DAG consists of a set of nodes V and a set $E \subset V \times V$ of ordered pairs such that if $(i,j) \in E$ then $(j,i) \notin E$. Groups of nodes in V form *paths* if for each two nodes i and j in the group, $(i,j) \in E$ or $(j,i) \in E$. Acyclic graphs have no cycles (paths with the same starting and ending node). Cycles lead to logical implausibilities (Pearl, 1986). Probabilistically, the joint distribution of the variables is not uniquely defined.

One of the fundamental relations expressible by a DAG is that of precedence. Parental nodes precede their children, in a causal or temporal sense, inducing an ordering of the nodes in the graph. We enumerate the nodes in a top–down fashion, starting with those without parents (source nodes) at the first level, proceeding to their children at the next level, and so on to the bottom of the graph. Nodes allocated to the same level can be ordered arbitrarily permitting many equivalent enumerations.

If node i precedes j, written $i \prec j$, according to an ordering in the graph, then $i \in pr(j) \equiv \{s : s \prec j\}$, the set predecessors of j. We also define the parental set of i as pa(i), and the set of children of i as ch(i) and denote the variable at node i by X_i. Conditional independence arises naturally for a DAG in terms of the set of predecessors. That is, if nodes i and j are not connected and $i \prec j$, then

$$j \perp\!\!\!\perp i \mid pr(j-i) <=> j \perp\!\!\!\perp i \mid pa(j)$$

where pr(j–i) is the set of predecessors of j excluding i. This is equivalent to saying that j is conditionally independent of i given its parents.

As a result, the following factorization of the joint density of the random variables at the nodes in V ensues (see e.g. Whittaker, 1990, p. 73). If n is the number of nodes in V,

$$f(x_1,\ldots,x_n) = \prod_{v=1}^{n} f(x_v \mid pa(v)) \qquad (1.2)$$

This factorization was illustrated in conjunction with figure 1 and is central to the stochastic simulation methodology described in section 2 as well as to the distribution theory in section 3.

2. Simulation Approaches

Desired analysis of general mixed directed graphical models requires high dimensional integration/summation. Such calculations usually can not be carried out analytically, either exactly or approximately. Simulation methods offer a viable alternative as shown in section 2.1. In particular, Monte Carlo methods may be performed noniteratively or iteratively. We detail these approaches in subsections 2.2 and 2.3 respectively.

2.1 Why Simulation?

The joint density of the set of variables $X = (X_1,X_2,\ldots,X_n)$ represented by the graphical model G is expressible by the factorization (1.2). The variables X_i may be either discrete or continuous. Explicit discussion of forms for the f's is deferred to section 3. Following earlier remarks we assume that, whatever form these specifications take, all the f's are fully known, i.e., included in the set $pa(X_i)$ are any parameters in the density of X_i. To answer questions we may be interested in regarding model G requires the ability to

compute arbitrary conditional distributions and their features. In particular, suppose the conditional information fixes a subset of the variables (possibly an empty set if marginalization is sought), X_0, of size n_0 to specified levels. The nodes in X_0 are called *evidence* nodes in Shachter and Peot (1989). Let X_u denote the complement of X_0, i.e. the *unobserved* nodes. We seek the conditional distribution of X_u given $X_0=x_0$ as well as that of X' given $X_0=x_0$, where X' denotes a generic component of X_u. Exact calculation of $f(X_u|X_0 = x_0)$ requires an $n - n_0$ dimensional integration/summation and calculation of $f(X'|X_0 = x_0)$ requires an additional $n_0 - 1$ dimensional integration/summation. Envisioning n and possibly n_0 to be large, such computation will not be feasible by exact or approximate analytic methods. Hence, we turn to simulation strategies.

As observed in Smith and Gelfand (1992), there is an essential duality between a sample and the density (distribution) from which it is generated. Clearly the density generates the sample. But conversely, given a sample we can approximately recreate the density and its features. Thus our objective is to draw samples from $f(X_u|X_0 = x_0)$. Drawing observations from the joint distribution of X is straightforward. It may be done in a "top down" fashion using the components of the factorization (1.2). That is, we sample all source nodes, then sample all their children, etc. The directed graphical model reveals which sampling orders are thus *feasible* and we shall in fact assume that the labeling of the X's from X_1 to X_n constitutes a feasible sampling order.

In attempting to sample $f(X_u|X_0 = x_0)$ one might take draws from f(X) and retain those meeting the evidence $X_0 = x_0$. Such rejection sampling is called logic sampling (see, e.g., Henrion 1988), and is very inefficient when the nodes in X_0 are discrete but the event $X_0 = x_0$ is rare; it is impossible if any of the nodes in X_0 are continuous.

Note that, though we can not obtain $f(X_u|X_0 = x_0)$ explicitly, we do know its form modulo a normalizing constant, i.e.,

$$f(X_u|X_0 = x_0) \propto f(X_u, x_0) \tag{2.1}$$

where the right hand side of (2.1) is given in (1.2). In fact we can write

$$f(X_u|X_o=x_o) \propto \prod_{X_i \in X_o} f(X_i|pa(X_i)) \prod_{X_i \in X_u} f(X_i|pa(X_i))\Big|_{(X_u,x_o)} \tag{2.2}$$

Suppose, as is traditional, that we refer to what is observed as "data" and what is unobserved as "parameters". Then the first product on the right side of (2.2) could be considered as a *likelihood* since here all the X_i are observed while the second product could be considered as a *prior* since here none of the X_i are observed. Then (2.2) is of the form Likelihood x Prior as in customary Bayesian modeling and we may bring to bear on our problem all of the recently discussed sampling–based technology for Bayesian calculations. This work includes noniterative Monte Carlo using importance sampling, as in Van Dijk and Kloek (1983), Stewart (1983), and perhaps best summarized in Geweke (1989) as well as iterative or Markov chain Monte Carlo as in Geman and Geman (1984), Tanner and Wong (1987) and Gelfand and Smith (1990) and perhaps best summarized in Tierney (1991).

2.2 Independent or Noniterative Monte Carlo

Since we can not sample from $f(X_u|X_o = x_o)$ directly, we develop and employ a suitable importance sampling density (ISD). More precisely from a density say $g(X_u)$ we draw a large sample denoted by X_{ut}, $t = 1,...,m$ where $g(X_u)$ has the same support as $f(X_u|X_o = x_o)$. Then, expectations under $f(X_u|X_o = x_o)$ e.g. $E(h(X_u)|X_o = x_o)$ are approximated by the weighted sum

$$\hat{h}_m = \frac{\sum_{t=1}^{m} h(X_{ut}) \cdot w_t}{\sum_{t=1}^{m} w_t} \tag{2.3}$$

where $w_t = f(X_{ut}|X_o = x_o)/g(X_{ut})$. Moreover, resampling the X_{ut} with probabilities $g_t = w_t/\Sigma w_t$ produces samples approximately distributed according to $f(X_u|X_o=x_o)$ (see Rubin, 1988, Smith and Gelfand, 1992).

The selection of g becomes the primary concern. The more closely g resembles

$f(X_u, x_o)$ the more efficient g is in terms of sample size m. Hence a good ISD is characterized as having fairly constant weights w_t. Such stability might be measured through the variance of the w_t (Geweke, 1989) or their entropy (Ferguson, 1983). A natural choice for g would be the *prior* $\prod_{X_i \in X_u} f(X_i | pa(X_i))\Big|_{(X_u, x_o)}$ which, provided that it is proper, it can be readily sampled using a feasible sampling order. This choice of g results in weights $w_t = \prod_{X_i \in X_o} f(X_i | pa(X_i))\Big|_{(X_{ut}, x_o)}$ which would naturally be called likelihood weights (see Shachter and Peot, 1989), i.e., bigger weights are attached to "more likely" X_{ut}'s. Such w_t are not likely to be very stable since we would not expect the prior to "match" the Likelihood x Prior form. Nonetheless, such weights are used successfully in the example of Section 4.

A more refined selection of g can be obtained as follows. Partition X_u into a set of discrete and a set of continuous variables denoted by X_u^d and X_u^c respectively. We consider an ISD which samples equally likely over the domain of X_u^d and then, given $X_u^d = x_u^d$, draws $X_u^c \sim g(X_u^c | X_u^d = x_u^d)$. The joint density of X_u under this ISD is proportional to the conditional density $g(X_u^c | X_u^d)$. Thus to match $f(X_u, x_o)$, for each $X_u^d = x_u^d$ we would choose $g(X_u^c | x_u^d)$ to match $f(X_u^c, x_u^d, x_o)$. Methods for developing an efficient ISD for a nonstandardized continuous density have received much attention lately (see e.g. Geweke, 1989, Oh & Berger, 1992, West, 1992). In most of this work the resultant ISD is a mixture of normal or t distributions.

2.2.1 Performance of Independent Monte Carlo Estimates

Results in Geweke (1989) applied to $\hat{h}_m$ of (2.3), show that under mild conditions,

(i) $\sqrt{m}[\hat{h}_m - Eh(X_u | X_o = x_o)]$ is asymptotically normal $N(0, \sigma^2)$, where $\sigma^2 = c^{-2}(x_o) \int \left[(h(x_u) - Eh(X_u | X_o = x_o))^2 \cdot f^2(x_u, x_o)/g(x_u) \right] dx_u$. Here $c(x_o) = \int f(x_u, x_o) dx_u$ with integral signs here and in the sequel denoting a multiple

integration/summation over the domain of X_o.

(ii) $\hat{\sigma}_m^2 = \sum_{t=1}^{m} [h(X_{ut}) - \hat{h}_m]^2 w_t^2 / (\sum_{t=1}^{m} w_t)^2$ which can be calculated from the sample of X_{ut} is such that $m\hat{\sigma}_m^2 \to \sigma^2$ enabling an estimate of the accuracy of $\hat{h}_m$.

Suppose we seek density estimates for the components of X_u perhaps to obtain modal values or quantiles. The X_{ut} can be resampled as suggested after (2.3), to create samples approximately distributed according to $f(X_u|X_o)$. Then, using appropriate coordinates, histograms for the X_u^d and kernel density estimates for the X_u^c can be created. Improved estimates utilizing the known forms $f(X_u,x_o)$ and $g(X_u)$ are available. If we want an estimate of $f(X'|X_o=x_o)$, where X' is a generic component of X_u, we may write,

$$f(X'|X_o=x_o)=\int \frac{f(X',x_u',x_o)}{g(x_u')} g(x_u')dx_u' \doteq \int \frac{f(x',x_u',x_o)}{g(x',x_u)} g(x',x_u')dx'dx_u'$$

where X_u' is the collection of all variables in X_u apart from X' and $g(X_u') = \int g(x_u)dx'$. Suppose $x_{ut}=(x_t',x_{ut}')$, t=1,2, ...,m, are generated from $g(X_u)$. Let

$$\hat{f}(X'|X_o=x_o) = \frac{\Sigma h_t(X')w_t}{\Sigma w_t} \tag{2.4}$$

where $h_t(X') = \frac{w_t(X')}{w_t}$, $w_t(X') = \frac{f(X',X_t',x_o)}{f(X_t')}$, and $w_t = \frac{f(X_t',X_{ut}',x_o)}{g(X_t',X_{ut})}$. Since (2.4) is of the form (2.3) results of Geweke apply. In particular, at any fixed X', we may estimate the accuracy of our density estimator using

$$\hat{\sigma}^2(X') = \sum_{t=1}^{m} [h_t(X') - \hat{f}(X'|X_o=x_o)]^2 \cdot w_t^2 / (\sum_{t=1}^{m} w_t)^2 \tag{2.5}$$

Note that w_t need not be recomputed with changing X'. However in computing $w_t(X')$, a univariate integration or summation over X' is required to obtain $g(X_u')$. In the continuous case, the integration can be done quickly and cheaply using a trapezoidal or Simpson's rule integration. The estimator in (2.4) is used in the illustrative example of section 4.

2.3 Dependent or Iterative Monte Carlo

Rather than using independent samples as in the ISD approach of section 2.2 we may generate dependent sequences using a Markov chain whose equilibrium distribution is $f(\underset{\sim}{X}_u | \underset{\sim}{X}_o = \underset{\sim}{x}_o)$. Such Markov chain Monte Carlo dates at least to Metropolis et al (1953). A version which samples from updated complete conditional distributions was introduced as the Gibbs sampler by Geman and Geman (1984) for restoration of noisy images. The setting is a high dimensional Markov random field with binary nodes built from local neighborhood structure. Pearl (1987) applied the Gibbs sampler (referring to it only as a stochastic simulation technique) to causal models involving binary variables. Gelfand and Smith (1990, 1991) discuss a version involving continuous nodes in the context of hierarchical Bayesian models.

The attraction of the Gibbs sampler is that draws from the high dimension density $f(\mathbf{X}_u | \mathbf{X}_o = \mathbf{x}_o)$ are obtained by making draws from the univariate complete conditional distributions $f(X' | \mathbf{X}'_u, \mathbf{x}_o)$. Given a starting state vector for $\mathbf{X}_u$, the components of $\mathbf{X}_u$ are typically sampled in the natural order, sometimes referred to as a raster scan, with the conditional levels of $\mathbf{X}'_u$ updated after each sampling while $\mathbf{X}_o$ remains "clamped" at $\mathbf{x}_o$ (In fact, any infinitely often visiting order of the components works). One pass through all of the components of $\mathbf{X}_u$ is called an iteration. After l such iterations a sampled vector $\mathbf{X}_u^{(l)}$ will result. Under conditions which insure that the iterations $\mathbf{X}_u^{(l)}$ are not trapped in a portion of the state space of $\mathbf{X}_u$, as $l \to \infty$, $\mathbf{X}_u^{(l)}$ converges in distribution to an observation from $f(\mathbf{X}_u | \mathbf{X}_0 = \mathbf{x}_o)$. Such conditions mandate that we can not permit any purely deterministic nodes in our graphical model. Technically, we merely remove any such nodes, adjusting parents and children accordingly. Convergence will be at a geometric rate, possibly uniformly so (see e.g. Roberts and Polson, 1990, Schervish and Carlin, 1991, Liu, Wong and Kong, 1991 and rather generally, Nummelin, 1984). Gelman and Rubin (1992) show that, for certain models, l will have to be extremely large before "acceptable" convergence is achieved and that several different starts for $\mathbf{X}_u$ will be required to make a

convergence assessment. An attractive empirical tool for assessing convergence is the Gibbs stopper as described in Tanner (1991). Unfortunately, apart from special cases, its computation becomes prohibitive with increasing n. However, use of subgraph approximation results in a substantially reduced graph, assisting in the examination of convergence.

It is worthwhile to remark that errors due to the quality of our various approximations will likely be small compared to the general uncertainty in the model specification itself. For complex high dimensional models incorporating arbitrary dependence structure, "ballpark" density estimates may have to suffice.

Let us be more specific about the form of the complete conditional distribution for X'. It is clear that $f(X'|X'_u, X_0=x_0)$ is proportional to (1.2). Moreover, with regard to factorization, only terms involving X' as a child or as a parent need be considered so that

$$f(X'|X'_u, X_0=x_0) \propto \left[f[X'|pa(X')] \prod_{V \in ch(X)} f[V|pa(V)] \right] \Big|_{(X_u, x_0)} \qquad (2.6)$$

Thus, the only variables involved in the complete conditional distribution of X', are its parents, its children, and the parents of these children. This set has been called the Markov blanket by Pearl (1986). Since typically dependence in the model is sparse, only a few of the terms in (1.2) appear in (2.6).

Turning to the sampling itself we recall that by virtue of the Markovian updating the conditional levels in (2.6) will change with each draw of X'. In addition, the form of (2.6) will almost never be a standard distribution even if the individual terms in the product are. For discrete variables sampling is routine upon standardization/summing of (2.6) although for ordinal variables rejection methods are available (Devroye, 1986). For continuous variables we might also consider a rejection method (see Devroye, 1986 or Ripley 1987). For instance an envelope function for (2.6) is $Mf[X'|pa(X')]$ where $M = \sup_{X'} \prod_{V \in ch(Y)} f(V|pa(V)) \Big|_{(X_u, x_0)}$. Typically, $f[X'|pa(X')]$ is readily sampled and M is not difficult to compute though it must be recomputed with changing levels of X'_u. In

practice such envelopes tend to be very inefficient producing very small acceptance rates. This is not surprising since the concentration of mass for $f[X' | pa(X')]$ may be quite different from that of (2.6). Nonetheless, we used such rejection for our example in Section 4 since generation from the univariate $f(X' | pa(X'))$ is so inexpensive.

Unfortunately, specialized rejection methods such as in Gilks and Wild (1991) which take advantage of log concavity of the density f, squeezing etc. will not likely be applicable to the form (2.6). Other tailored versions may be costly and not readily automated. Alternative "black box" Gibbs samplers for graphical models handle continuous variables by approximate inversion of the probability integral transform as in, e.g., Tanner (1991), by the use of a modified ratio of uniforms method as described in Wakefield, Gelfand and Smith (1991), by adaptive normal kernel density approximations to (2.6) (Silverman, 1986, §5.3).

As for density estimation utilizing the output of the Gibbs sampler, suppose we have a sample of m vectors X_{ut}, t=1,2,...,m, roughly independent and approximately distributed from $f(X_u | X_0 = x_0)$. For component X' we could create a histogram from the set X'_t for discrete X' or a kernel density estimate (as in Silverman, 1986) for X' continuous. However, Rao–Blackwellized density estimates,

$$\hat{f}(X' | X_0 = x_0) = \frac{1}{m} \sum_{t=1}^{m} f(X' | \underset{\sim}{X}'_{ut}, \underset{\sim}{X}_0 = \underset{\sim}{x}_0), \tag{2.7}$$

using the known structure (2.6), are preferable (see in particular, Gelfand and Smith 1991).

3. Distributional Models and Model Choice

Thus far we have assumed complete specification of the directed graphical model. As emphasized earlier the attraction of these models is that the global structure is determined through provision of local detail. Specification of local detail is at the node level in accordance with the factorization (1.2) and encompasses two aspects – identification of parents and of conditional distributions at the node given levels of the

parental variables.

In models comprised of an ensemble of variables with, a priori, little structural insight, often the purpose of the modeling exercise is to identify the nature of the dependence. In such contexts edges of the graph need not be directed so that there can be links between nodes without labeling a parent and a child. The book of Whitaker (1990) provides a good account of the process of modeling dependence. See also the recent papers of Wermuth and Lauritsen (1990) and Edwards (1990). Our models are directed, i.e., restricted to systems having a natural flow or hierarchical sequence for the variables whence identification of parents is implicit. It is unrealistic to assume that the dependence structure is known so precisely. For a given set of nodes we might postulate several structural forms i.e. several directed graphical models, and attempt to choose amongst them. We return to this issue at the end of the section.

As for distributional specification, in practice such distributions should be elicited. The Bayesian literature on elicitation may be useful in this regard. See, for example, Kahneman et al (1982) for a very readable discussion, and Kadane et al (1980) for implementation suggestions. Still, several functional forms may be candidate distributions again leading to the question of model choice. An attractive feature of the simulation methodology developed in section 2 is that it can be applied regardless of the forms taken for the $f(X_i|pa(X_i))$.

Classes of models for discrete nodes are discussed in Spiegelhalter and Lauritzen (1990). In particular, for an arbitrary node Y let us think of pa(Y) as comprised of two sets, one a set of "causal" parents, γ_Y and the other a set of "parametric" parents, θ_Y. Then X is viewed as the collection of nodes (V,θ) where V is the set of all "nonparameter" nodes and $\theta = \bigcup_{Y\in V} \theta_Y$ provides the overall parametrization for the model. Then (1.2) becomes

$$f(V, \theta) = \prod_{V_i \in V} f(V_i|\gamma_{V_i}, \theta_{V_i}) \cdot \prod_{\theta_i \in \theta} f(\theta_i|pa(\theta_i))$$

Customary modeling imposes $pa(\theta_i) \subset \theta$.

Spiegelhalter and Lauritzen introduce an assumption of *global independence* which says that $\prod_{\theta_i \in \theta} f(\theta_i | pa(\theta_i)) = \prod_{V_i \in V} f(\theta_{V_i})$. Thus the θ_{V_i} are presumed to be independent sets of parameters. If all of the parents of Y are discrete variables we can envision θ_Y as breaking into components each corresponding to a configuration of pa(Y). Spiegelhalter and Lauritzen then introduce the further assumption of *local independence* if $f(\theta_Y)$ factors into a product of densities over these components.

Now if Y is a nominal variable its realizations constitute multinomial sampling in which case the component of θ_Y corresponding to a particular configuration of pa(Y) could be taken as the vector of multinomial probabilities for Y under this configuration. A convenient choice of density for this component would be a Dirichlet whence, under local independence, $f(\theta_Y)$ would be a product of Dirichlet densities over all possible onfigurations of pa(Y). As Spiegelhalter and Lauritzen observe, such specification introduces, for Y alone, θ_Y of dimension equal to (# of levels of $Y - 1$) x (# of configurations of pa(Y)). A more parsimonious assumption, which also accommodates the case where Y has at least one continuous parent, models $f(Y|\gamma_Y,\theta_Y)$ as a parametric family in θ_Y with γ_Y as covariates. For instance we might model the multinomial probabilities for Y through a set of logits (e.g., baseline or adjacent). In the binary case we would set

$$\log\left(\frac{f(Y=1|\gamma_Y,\theta_Y)}{f(Y=0|\gamma_Y,\theta_Y)}\right) = T(\gamma_Y)^T \cdot \theta_Y$$

where $T(\gamma_Y)$ is an $r \times 1$ vector of functions of γ_Y and θ_Y is thus an $r \times 1$ vector of coefficient parameters. We might further assume a multivariate Gaussian density for θ_Y, possibly rather vague.

In the case of mixed models, the unique distributional family which has been discussed in the literature is the conditional Gaussian family (Lauritzen and Wermuth, 1989; Wermuth and Lauritzen, 1990). The conditional Gaussian distribution provides a

joint distribution for the entire directed graphical model. It arises from a two part asymmetric construction. The marginal distribution of the discrete variables is multinomial and, condition on these variables, the continuous variables have a multivariate normal distribution. The form of the joint density (1.2) is then immediate and is parametrized by the set $\{p_i,\mu_i,\Sigma_i\}$ where i indexes the state space of the set of discrete nodes and p_i is the probability of state i with μ_i and Σ_i the mean vector and covariance matrix respectively for the conditional multivariate normal distribution given state i. The conditional Gaussian family has an attractive set of theoretical properties which are summarized in Lauritzen (1990b) and in Whittaker (1990). Unfortunately, this family is restrictive in two unappealing ways. First, no discrete node may have a continuous parent and second, all continuous nodes must follow a normal distribution. However, using this family, exact calculations can be carried out using the junction tree representation as in Lauritzen (1990a). If not, Lauritzen (1990a) provides only an approximate computational method using second order Taylor series expansions which produces moment estimates for the conditional density $f(Y|X_0=x_0)$, but no density estimates.

We propose handling continuous nodes as we did the discrete nodes above, i.e., assuming $f(Y|\gamma_Y,\theta_Y)$ to be a parametric family in θ_Y with γ_Y as covariates. A broad and convenient class of specifications would be those of generalized linear models (McCullagh and Nelder, 1989) where $f(Y|\gamma_Y,\theta_Y)$ is a univariate continuous exponential family and a suitable link function connects the mean of Y with a linear form in θ_Y. Discrete nodes modeled as e.g., binomial or Poisson, could also be handled in this fashion.

We return to the matter of model choice. Consider two competing fully specified directed graphical models to describe a system. The models may differ in terms of parental specification at one or more nodes and/or distributional specification at one or more nodes. How do we choose between the models? The question is not well formulated unless we ask it with regard to data observed at components of the system. Suppose X_0 denotes the set of nodes at which we take observations and suppose we observe independent x_{0l}, $l = 1,...,L$.

Note that the x_{0l} may be real data generated from the system or artificial data generated under some joint distribution for X_0 which we would like the system to emulate. It seems natural to choose the model more likely to have yielded this data. Using the methodology in section 2, we can calculate under each model an estimate of the density $f(X_0)$, hence, an estimate of $\prod_l f(x_{0l})$. The model with the larger value of this product would be chosen. Unless the dimension of X_0 is small, a very large L may be required to obtain a satisfactory estimate of $f(X_0)$. Instead we might replace $f(X_0)$ by $\prod_{X' \in X_0} f(X' | X_0')$, where X_0' is the collection of all variables in X_0 apart from X'. The densities in this product are univariate and can be straightforwardly estimated using ideas in section 2. In particular, with X_u defined as before, suppose X_{ut}, t=1,2,...,m, is a sample from $f(x_u | x_0')$. Then

$$\hat{f}(x' | x_0') = m^{-1} \sum_{t=1}^{m} f(x' | X_{ut}, x_0') \qquad (3.1)$$

Note that (3.1) is a mixture of complete conditional distributions for X'. See Gelfand, Dey and Chang (1992) for further discussion of the use of such cross validated densities in the context of model selection.

4. An Illustrative Example

Accurate diagnosis of congenital heart disease immediately after birth increases the newborn's chances of survival but is a far from simple procedure. In the interest of communication of preliminary symptoms between admitting and specialist physicians and education of medical students, a questionnaire was developed to aid doctors during an infant's referral to the specialist hospital (Franklin et al., 1991, Spiegelhalter et al., 1991). A directed graph representing one aspect of congenital heart disease appears in figure 3.

[Insert figure 3 about here]

The graph and extensive conditional probability tables quantifying the associations between disease, symptoms, evidence and risk factors were constructed in consultation with a group of pediatric cardiologists, and used by Spiegelhalter and Cowell (1991) to describe learning about unobserved variables (in this particular case all internal nodes). The conditional probability tables were kindly provided by Dr Spiegelhalter to aid our analysis. In their analysis all nodes in the graph were assumed discrete. In fact, several of the nodes such as 11, 12, 16 and 17 are inherently continuous, but a CG family of models cannot permit the first two to of these to be so.

We analyze this graphical model by the Monte Carlo methods of Section 2. Logit forms were used to describe the conditional distributions at the discrete nodes, while the continuous nodes (11, 12, 16 and 17) were assumed to be normal on the logarithmic scale. The conditional normal distributions for the continuous nodes were chosen (i.e., mean and variance) to yield probabilities which essentially match the interval probabilities in the provided probability tables. Table 1 presents marginalization results, listing for both the Gibbs and the importance sampling method, estimates of marginal probabilities for the discrete nodes and estimates of the mean and variance for the continuous nodes. All estimates presented are based on 5000 random generations. Differences in the estimates achieved by these two methods are small and within anticipated variability based on the number of replications. Plots of the continuous marginal density estimates based on the same number of generations from both techniques are overlaid in figures 4a through 4d. Finally, to illustrate the revision of probabilities under given information, we update the probabilities for each of the diagnoses at node 3 given clinical data (levels of nodes 15 through 20). For instance if LVH was *reported*, RUQ O_2 = 1.5 (which corresponds to 4.48 units), Lower Body O_2 = 1.4 (4.05 units), reported CO_2>7.5, X–ray was *normal*, and grunting was *reported*, the probabilities for disease (node 3) become,

PFC	TGA	Fallot	PAIVS	TAPVD	Lung	
0.02833	0.17419	0.16339	0.53830	0.04844	0.04735	(GIBBS)
0.01622	0.17838	0.16103	0.52048	0.06454	0.05936	(Imp. Samp.)

and can be compared to the results in Table 1. The evidence point towards PAIVS and away from lung diseases (influenced by the report of a normal X–ray).

Table 1. Congenital Heart Disease Example; Initial State (no evidence). Estimates of probabilities of discrete variables and means and variances of continuous nodes as produced by the Gibbs sampler (first entry), and Importance Sampling (second entry), based on 5000 generations.

(1) Birth Asphyxia?

Yes	No
0.09859	0.90141
0.09830	0.90170

(2) Age

0–3 days	4–10 days	11–30 days
0.33098	0.33335	0.33568
0.33320	0.33300	0.33380

(3) Disease

PFC	TGA	Fallot	PAIVS	TAPVD	Lung
0.03816	0.23229	0.34924	0.20520	0.13819	0.03692
0.04120	0.24320	0.35340	0.20280	0.12700	0.03240

(4) Duct Flow

Left to Right	None	Right to Left
0.56197	0.32355	0.11449
0.55810	0.32750	0.11440

(5) Cardiac Mixing

None	Mild	Complete	Transp.
0.03868	0.12052	0.64088	0.19993
0.04340	0.12070	0.62480	0.21110

(6) Lung Parench

Normal	Congested	Abnormal
0.49429	0.25894	0.24677
0.50670	0.26170	0.23160

(7) Lung Flow

Normal	Low	High
0.18923	0.51253	0.29824
0.19260	0.51070	0.29670

Table 1. (Cont'd)

(8) Sick?

Yes	No
0.29949	0.70051
0.29160	0.70840

(9) LVH

Yes	No
0.25742	0.74258
0.25750	0.74250

(10) Hyp. Distribution

Equal	Unequal
0.50673	0.49327
0.50530	0.49470

(11) Hypoxia in O_2

Mean: 3.70686	Variance: 0.40038
Mean: 3.70287	Variance: 0.40199

(12) CO_2

Mean: 3.49466	Variance: 0.49334
Mean: 3.48318	Variance: 0.48438

(13) Chest X-ray

Normal	Oligaemic	Plethoric	Grd/Glass	Asy/Pathcy
0.13991	0.11527	0.15867	0.36094	0.22522
0.14540	0.12040	0.15320	0.36110	0.21990

(14) Grunting

Yes	No
0.36674	0.63272
0.36320	0.63680

(15) LVH Reported?

Yes	No
0.26828	0.73172
0.26620	0.73380

(16) RUQ O_2 (Continuous, see figure 5a).

Mean: 2.10503	Variance: 0.63272
Mean: 2.15679	Variance: 0.63136

(17) Lower Body O_2

Mean: 2.10503	Variance: 0.53596
Mean: 2.15617	Variance: 0.53570

Table 1. (Cont'd)

(18) CO_2 Report

<7.5	>=7.5
0.80591	0.19149
0.81530	0.18470

(19) X-ray Report

Normal	Oligaemic	Plethoric	Grd/Glass	Asy/Pathcy
0.18711	0.11816	0.19252	0.24481	0.25740
0.18710	0.11920	0.19558	0.24490	0.25300

(20) Grunting Reported?

Yes	No
0.35788	0.64212
0.35650	0.64350

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice (with discussion). *J. R. Statist. Society*, B, **36**, 192–326.

Darroch, J. N., Lauritzen, S. L. & Speed, T. P. (1980). Markov Fields and Log–Linear Interaction Models for Contingency Tables. *The Annals of Statistics*, **8**, 522–539.

Devroye, L.(1986). *Non–Uniform Random Number Generation.* Springer–Verlag, New York.

Edwards, D. (1990). Hierarchical Interaction Models. *J. R. Statist. Society*, B, **52**, 3–20.

Ferguson, T. (1983). Bayesian Density Estimation by Mixtures of Normal Distributions. *Recent Advances in Statistics*, M. Haseeb Rizvi et. al eds., Academic Press: New York, 287–302.

Franklin, R. C. G., Spiegelhalter, D. J., Macartney, F. J. & Bull, K. (1991). Evaluation of a Diagnostic Algorithm of Heart Disease in the Neonate. *British Medical Journal*, **302**, 935–939.

Gelfand, A. E., Dey D. & Chang, H. (1992). Model Determination Using Predictive Distributions With Implementation Via Sampling–Based Methods. In *Bayesian Statistics 4*. Eds. J. Bernardo et al., Oxford University Press (to appear).

Gelfand, A. E. & Smith A. F. M. (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398–409

Gelfand, A. E. & Smith A. F. M. (1991). Gibbs Sampling for Marginal Posterior Expectations. *Communications in Statistics, Theory and Methods*, **20**, 1747–1766.

Gelman, A. & Rubin, D. B. (1992). A Single Series from the Gibbs Sampler Provides a False Sense of Security. In *Bayesian Statistics 4*, Eds. J. Bernardo et al, Oxford University Press (to appear).

Geman, S. & Geman, D (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Simulation. *Econometrica*, **57**, 1317–1339.

Henrion, M. (1988). Propagating Uncertainty in Bayesian Networks by Probabilistic Logic Sampling. In *Uncertainty in Artificial Intelligence*, Vol. 2,J. F. Lemmer, and L. N. Kanal, (Eds.). North–Holland, Amsterdam, 149–164..

Jensen, F. V., Lauritzen, S. L., & Olesen K. G. (1991). Bayesian Updating Probabilistic Networks by Local Computations. *Computational Statistics Quarterly* (to appear).

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. & Peters, S. C. (1980). Interactive Evaluation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association*, **75**, 845–854.

Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Bias.* Cambridge University Press, New York.

Lauritzen, S. L. (1990a). Propagation of Probabilities, Means and Variances in Mixed Graphical Association Models. *Research Report* 90–18, Aalborg University.

Lauritzen, S. L. (1990b). Mixed Graphical Association Models, *Scandinavian Journal of Statistics*, 16, 273–306.

Lauritzen, S. L. & Wermuth, N. (1989). Graphical Models for Association between Variables Some of Which are Qualitative and Some Quantitative. *Annals of Statistics*, 17, 31–54.

Lauritzen, S. L. & Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems. *Journal of the Royal Statistical Society*, series B , 50, 57–224.

Liu, J. ,Wong, W. & Kong, A. (1991). Correlation Structure and Convergence Rate of the Gibbs Sampler (I): Applications to the Comparison of Estimators and Augmentation Schemes, Tech. Report 299, Dep't of Statistics, University of Chicago.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall: London

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087–1091.

Nummelin, E. (1984) *General Irreducible Markov Chains and Non–Negative Equations.* Cambridge University Press, Cambridge.

Oh, M–S, & Berger, J. O. (1992). Adaptive Imputation Sampling in Monte Carlo Integration. *Journal of Statistical Computing and Simulation* (to appear).

Pearl, J. (1986). Propagation and Structuring in Belief Networks. *Artificial Intelligence*, 29, 241–288.

Pearl, J. (1987). Evidential Reasoning Using Stochastic Simulation of Causal Models. *Artificial Intelligence*, 32, 247–257.

Quenouille, M. H., (1956). Notes on bias in estimation. *Biometrika*, 43, 353–360.

Ripley, B. (1987). *Stochastic Simulation.* John Wiley and Sons, New York.

Roberts, G. O. (1991). Convergence Diagnostics of the Gibbs Sampler. *Research Report*, 91–11, *Dept. of Mathematics, University of Nottingham*

Rubin, D. (1988). Using the SIR Algorithm to Simulate Posterior Distributions. In *Bayesian Statistics 3.* Eds. J. Bernardo et al., Oxford University Press, 395–402.

Schervish, M. and Carlin, B. (1991). On the Convergence of Successive Substitution Sampling, *Journal of Computation and Graphical Statistics* (to appear).

Shachter, R. D.& Peot, M. A. (1989). Evidential Reasoning Using Likelihood Weighting. *Artificial Intelligence*, (submitted).

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall: London.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian Statistics without tears: a sampling–resampling perspective. *The American Statistician*, (to appear).

Spiegelhalter, D. J., and Lauritzen, S. L. (1990). Techniques for Bayesian Analysis in Expert Systems. *Annals of Mathematics and Artificial Intelligence*, 2.

Spiegelhalter, D. J. and Cowell, R. G. (1991). Learning in Probabilistic Expert Systems. Presented in during the fourth Valencia international meeting on bayesian statistics.

Spiegelhalter, D. J., Harris, N. L., Bull, K. L. and Franklin, R. C. G. (1991). Empirical Evaluation of Prior Beliefs about Frequencies: Methodology and a Case Study in Congenital Heart Disease. *Research Report* 91–4, MRC Biostatistics Unit,. Cambridge, U.K..

Stewart, L. (1983). Bayesian Analysis using Monte Carlo Integration. A powerful Methodology for Handling Some Difficult Problems. *The Statistician*, 32, 195–200.

Tanner, M. (1991). *Tools for Statistical Inference: Observed Data and Data Augmentation Methods.* Lecture Notes In Statistics, Springer–Verlag, New York.

Tanner, M. and Wong, W. (1987). The Calculation of Posterior Distributions by Data Augmentation (with discussion). *Journal of the American StatisticalAssociation*, 82, 528–550.

Tierney, L. (1991). Markov Chains for Exploring Posterior Distributions. *Technical Report*, 560, School of Statistics, Univ. of Minnesota.

Thomas, A., Spiegelhalter, D. J., and Gilks, W. (1991). BUGS: A Program to Perform Bayesian Inference Using Gibbs Sampling. *MRC Biostatistics Unit Tech. Report*, Cambridge, England.

Van Dijk, H. and Kloek, T. (1983). Experiment with some Alternatives for Simple Importance Sampling in Monte Carlo Integration. In Bayesian Statistics 2 J. Bernardo et al.(Eds.), 511–530.

Wakefield, J., Gelfand, A. E., and Smith, A. F. M. (1992). Efficient Computation of Random Variates via the Ratio–of–Uniforms Method. Statistics and Computing, 1, 129–134.

Wermuth, N. and Lauritzen, S. L. (1990). On Substantive Research Hypothesis, Conditional Independence Graphs and Graphical Chain Models. *J. R. Statist. Soc.*, B, 52, No. 1, 21–50.

West, M. (1992). Approximating Posterior Distributions by Mixtures. In *Bayesian Statistics 4*, Eds. J. Bernardo et al, Oxford University Press (to appear).

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* John Wiley and sons, New York.

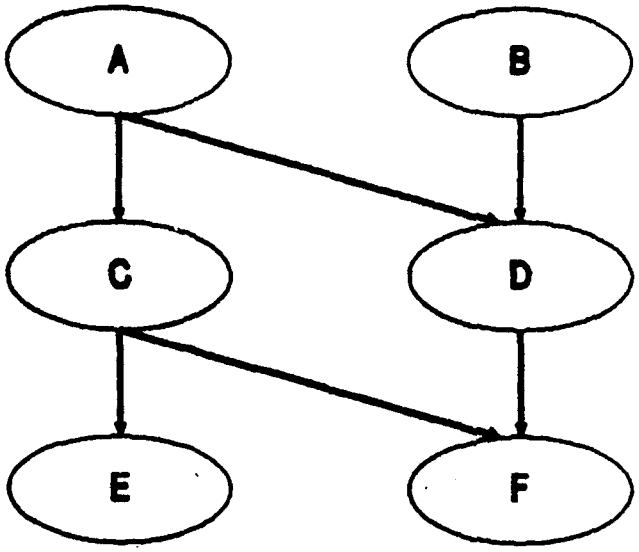

figure 1. A six node graphical model representing the joint distribution f(a,b,c,d,e,f) = f(e|c)f(f|c,d)f(c|a)f(d|a,b)f(a)f(b).

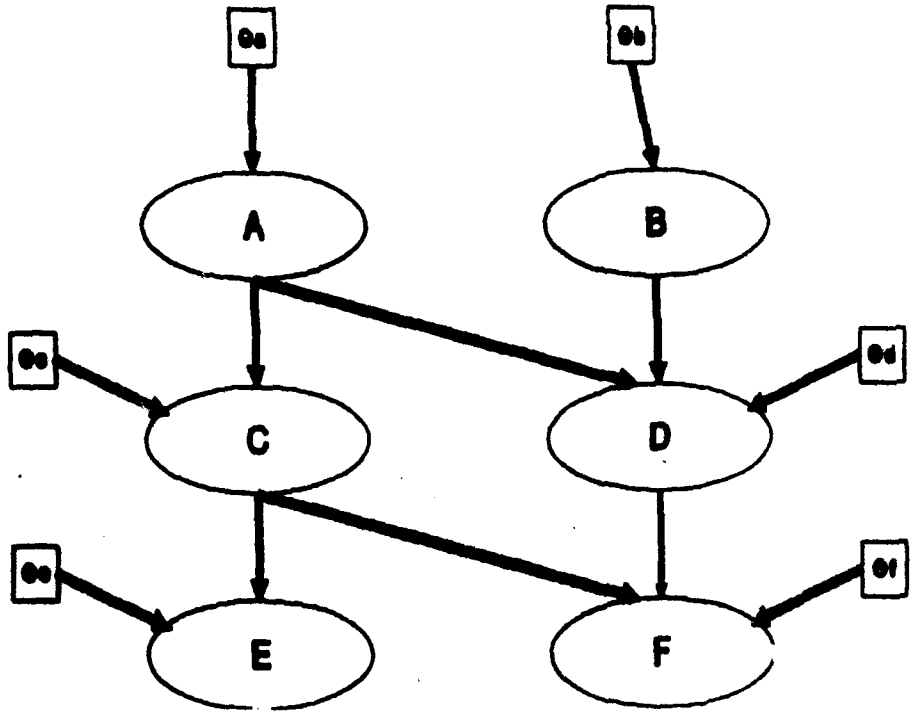

figure 8. Introduction of parameters in the graph of figure 1 to model uncertainty. Thus $f(a,b,c,d,e,f) = \int f(e|c,\theta_e) f(f|c,d,\theta_f) f(c|a,\theta_c) f(d|a,b,\theta_d) f(a|\theta_a) f(b|\theta_b) f(\theta) d\theta$

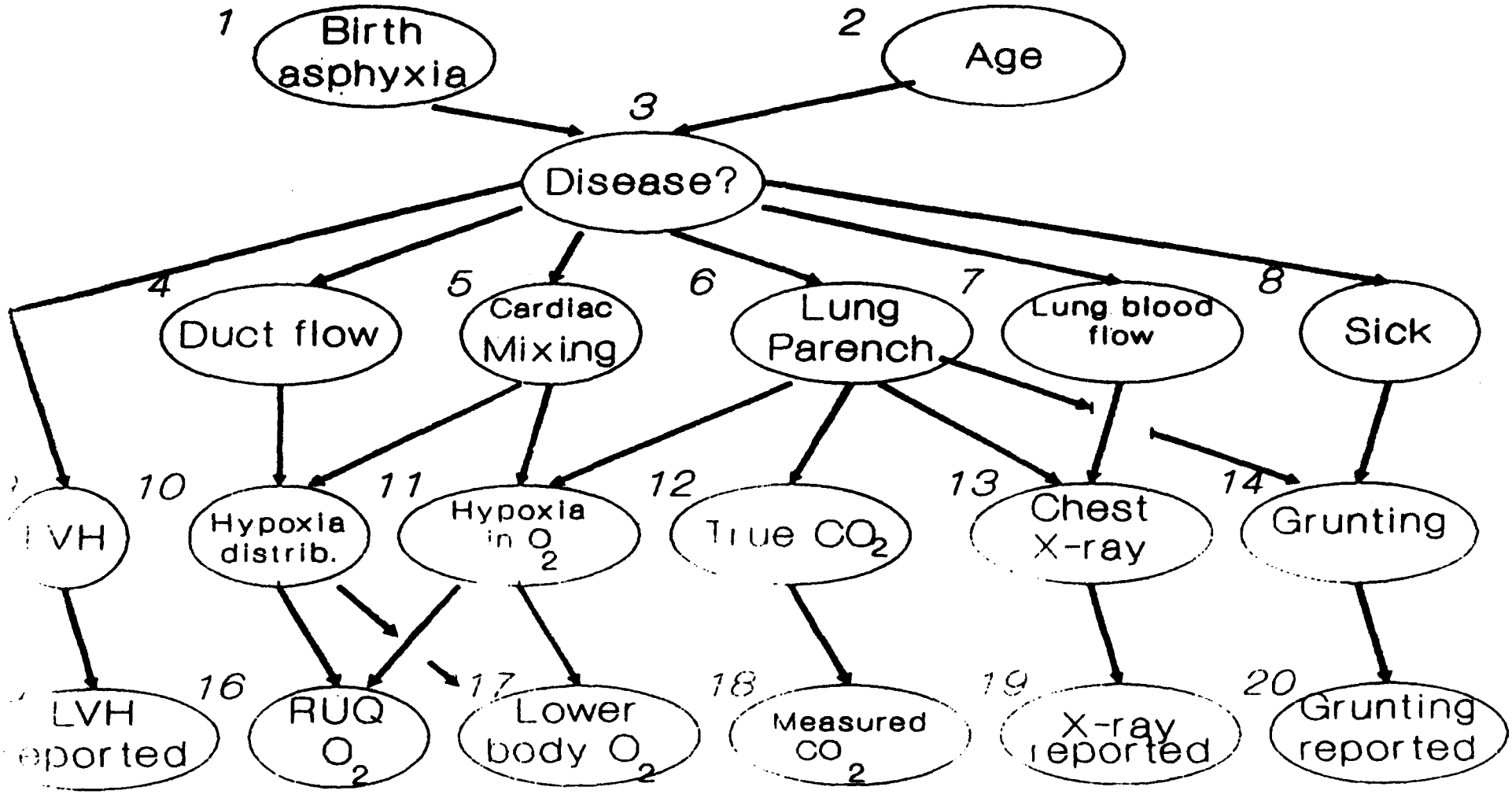

figure 3. **A directed graphical model for diagnosis of heart disease in neonates.**

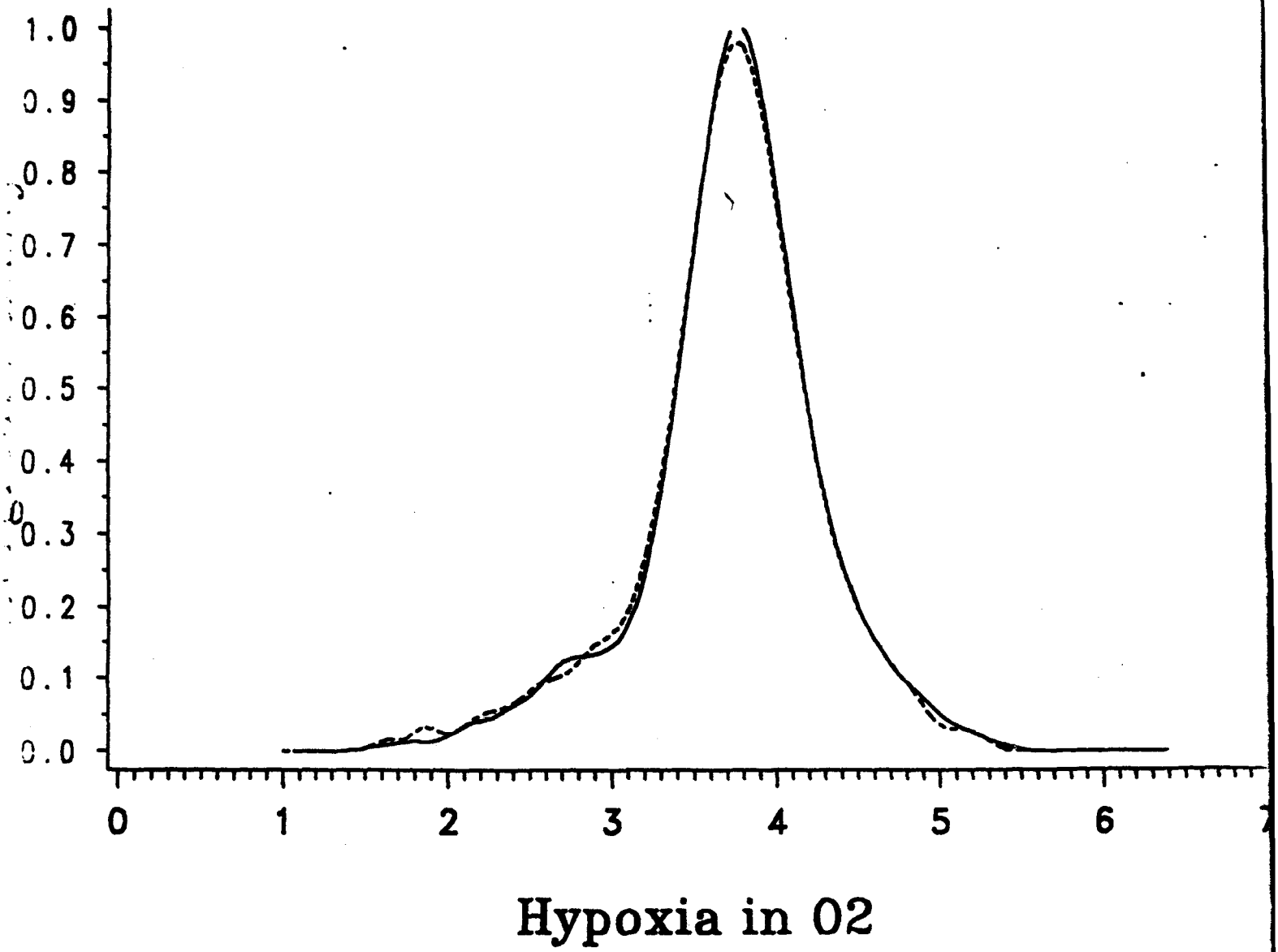

figure 4a. **Hypoxia in O_2. Plot of marginal density estimates using Gibbs and importance sampling, m = 1000 replications.**

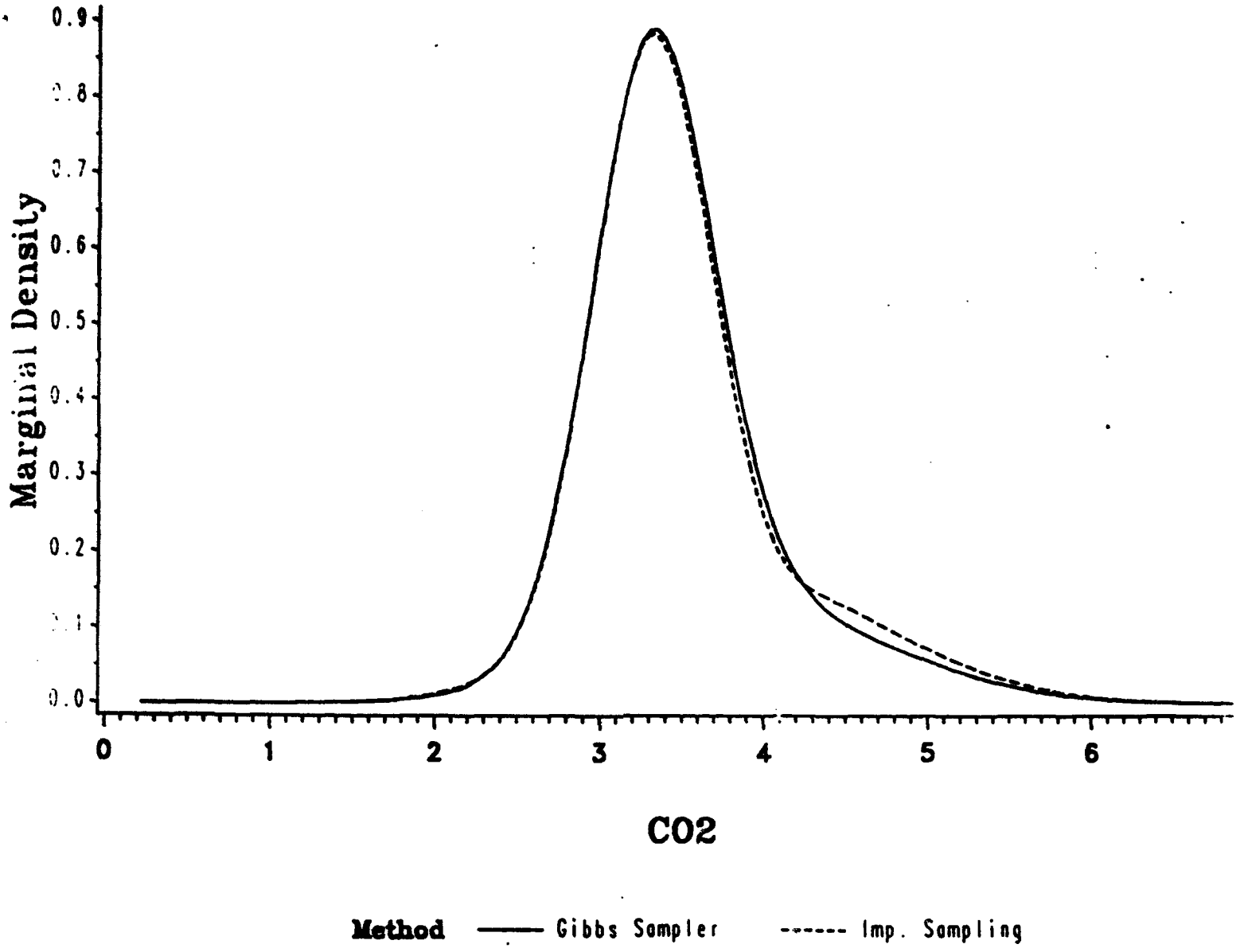

figure 4b. CO_2. Plot of marginal density estimates using Gibbs and importance sampling, m = 1000 replications.

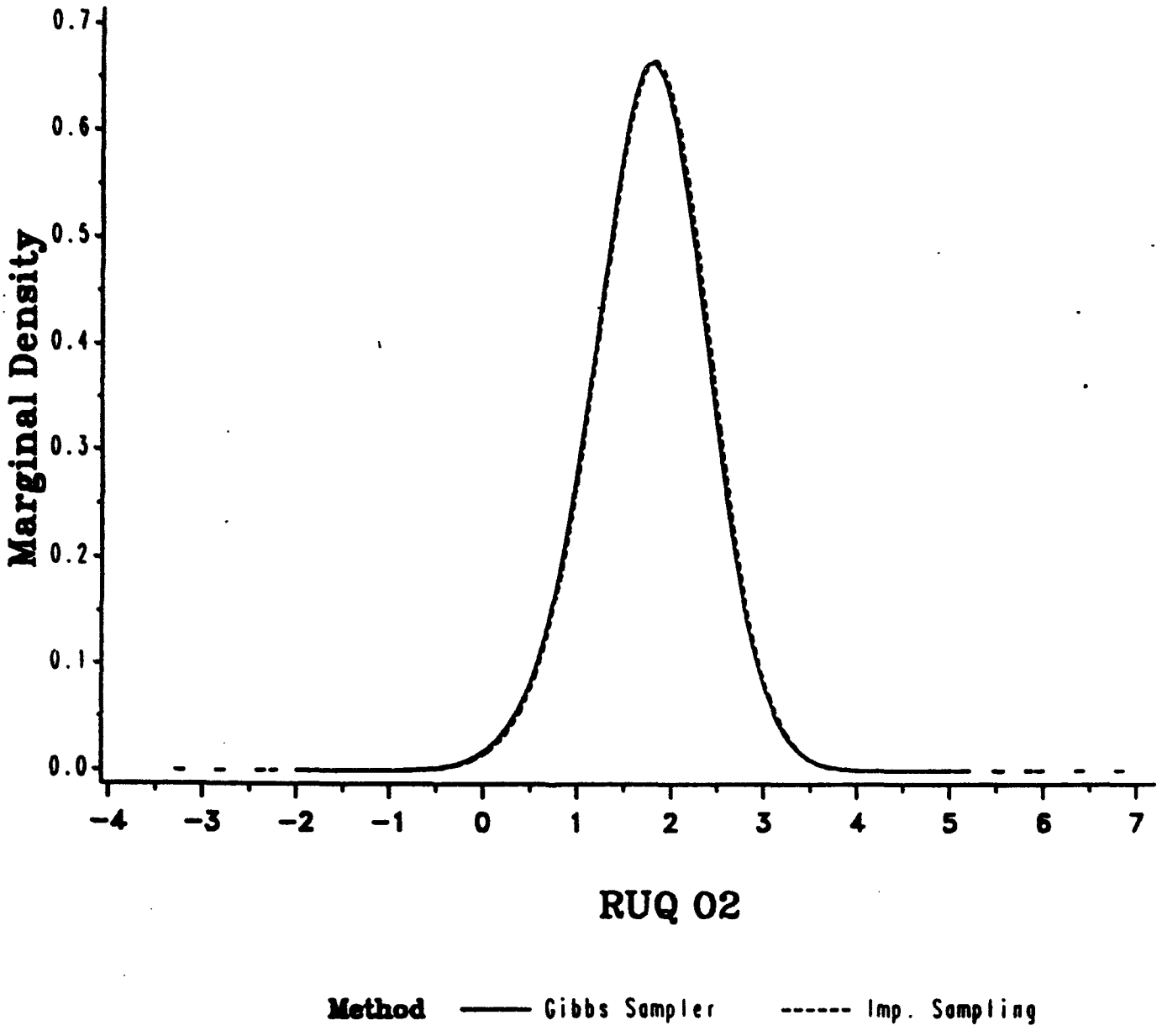

figure 4c. RUQ O_2. Plot of marginal density estimates using Gibbs and importance sampling, m = 1000 replications.

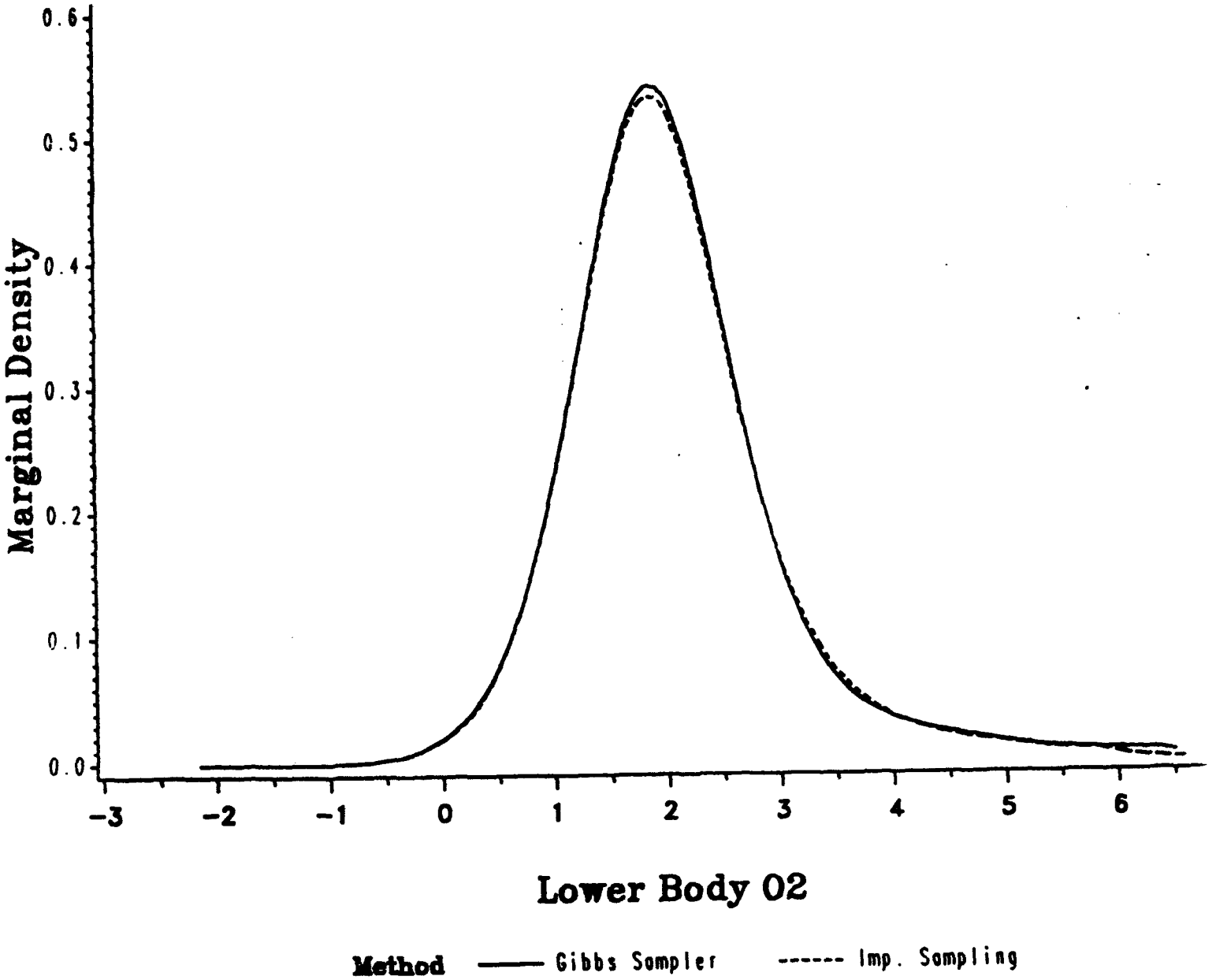

figure 4d. Lower Body O_2. Plot of marginal density estimates usin$_c$ Gibbs and importance sampling, m = 1000 replications.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ DISTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 475	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Simulation Approaches for Calculations in Directed Graphical Models		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Constantin T. Yiannoutsos and Alan E. Gelfand		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE October 12, 1993 13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Conditional independence, Gibbs sampler, likelihood weighting, Monte Carlo.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

ABSTRACT

In formulating models for a complex system graphical representation is an effective tool. When the components of the system are viewed as random variables, directed graphical models detail the nature of the dependence among them. Moreover, if for each variable the conditional distribution is provided according to the graph, the joint distribution is uniquely determined. Natural questions arise about the static behavior of the system under such specification as well as its response to information (observed levels of some of the variables). Answers to these questions require the ability to calculate arbitrary marginal and conditional distributions. In complex cases (high dimensional structures) such calculations require high dimensional integrations and/or summations. Most of the work to date has taken advantage of properties of directed graphs to facilitate exact calculations but is limited with regard to distributional assumptions and feasible system size. Monte Carlo methods for such calculations can accommodate much larger system size with arbitrary dependence structure and distributional forms yielding approximations which can be as accurate as desired. It is the objective of this paper to detail such methodology. An illustration is provided using a diagnostic system for congenital heart disease in neonates.