Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

NUMERICAL TECHNIQUES FOR SOLVING ESTIMATION PROBLEMS ON ROBUST BAYESIAN NETWORKS[1]

BY A. GAIVORONSKI, M. MORASSUTTO, S. SILANI AND F. STELLA
Italtel and Universitá di Milano

The Bayesian net is a formalism for structuring multidimensional distribution when initial data are scarse. It proved to be very useful in modeling of systems which depend on random parameters, in particular in image processing, reliability analysis, processing of medical information. In these cases very often it is impossible to recover the distribution with reasonable precision, but it is possible to identify a set of distributions to which the true distribution belongs. We consider the problem to define the lower and the upper bound for the functional defined on such a set. This gives rise to nontrivial optimization problems in the space of probability measures. We describe some algorithms for solving such optimization problems based on random search and linear programming techniques.

1. Introduction. In this paper we present numerical algorithms for solving problems of a special type which arise in a priori and a posteriori estimation of functions defined on Bayesian Networks (Pearl (1988)). In particular we present some results on modelling and optimization of complex stochastic systems in the case when the distribution functions of random parameters are only partially known (Ermoliev *et. al* (1985), Gaivoronski (1986)).

We consider systems which can be described by means of a set of functions $f^k(x,\zeta) : X \times \Omega \to \Re, k = 0,1,\ldots n$, where $x \in X \subseteq \Re^q$ represents controlled parameters and ζ is the vector of random parameters defined on appropriate probability space.

Usually one is interested in estimate of some characteristics of a system for fixed values of control parameters x, i.e. in finding the estimates of

$$\mathbf{E}\, f^k(x,\zeta) = \int_\Omega f^k(x,\omega)\, \mathrm{d}H^*(\omega), \quad k = 0,1,\ldots n \tag{1}$$

where ω and H^* are respectively a realization and the distribution of random parameters ζ. The next and more difficult step is to select the control parameters x in an optimal way, i.e. solve the following optimization problem:

$$\min_{x \in X} \mathbf{E}\, f^0(x,\zeta) \tag{2}$$

with possible constraints on values of functions $\mathbf{E}\, f^k(x,\zeta)$.
In this paper we consider the case when the distribution function H^* is not

[1] AMS 1991 *Subject Classifications.* Primary: 30E05, secondary: 65K10.
Key words and phrases. Generalized moment constraints, Bayesian Networks, estimations procedures, manufacturing.

completely known due to large dimension of the vector of random parameters ζ and/or scarsity of the experimental data. This situation is common in image processing (Geman *et. al* (1984)), reliability analysis (Gaivoronski *et. al* (1994)), processing of medical data (Lauritzen *et. al* (1993)) and many econometric problems. In this case it is not possible to solve the problems (1), (2) and new approaches are needed. One such possible approach is the following.

1. Structural Analysis. Clarify the structure of the random vector ζ. In our case this means finding statistical dependencies between various components of ζ in such a way that the joint distribution ζ can be represented as follows:

$$H^*(\zeta) = \prod_{v \in V} H^*(\zeta_v | \zeta_{c(v)}) \tag{3}$$

For example the set $c(v)$ can be a subset of vertices in V. Notice that in order for this representation to be correct it is necessary that conditional distirbutions $H(\zeta_V \mid \zeta c(v))$ satisfy compatibility constraints. Distributions of the type (3) are associated with the so-called *Bayesian Nets* (Pearl (1988)). More detailed information about this problem will be given in Section 2.

2. Robustness Analysis. Even after structural analysis very often the statistical evidence is not sufficient to define unique distributions $H(\zeta_v|\zeta_{c(v)})$. In this case it is possible to define for each pair $(v, c(v))$ the set $G_{v,c(v)}$ such that

$$H(\zeta_v|\zeta_{c(v)}) \in G_{v,c(v)}, \quad v \in V \tag{4}$$

These sets can be defined using experimental data and/or expert estimates. For example, they can be defined by bounds on mean, variance and other moments of $H(\zeta_v|\zeta_{c(v)})$ and/or bounds on probabilities of some sets by using *generalized moment constraints* (Dall'Aglio *et. al* (1994)) of the form:

$$\int_\Omega f^k(x, \omega_V)\, dH(\omega_V) \leq 0 \qquad k = 1, 2, \ldots n. \tag{5}$$

Combining the structure (3) with system of sets (4) we obtain *Robust Bayesian nets.* We refer to the notion of robustness here because now we can redefine the problems (1), (2) in such a way that their solutions will be robust with respect to all distributions defined by (3) and (4).

In particular, the solution of problem

$$\max_{H(\zeta_v|\zeta_{c(v)}) \in G_{v,c(v)}} \int_\Omega f^k(x, \omega_V) \prod_{v \in V} \mathrm{d}H(\omega_v|\omega_{c(v)}) \tag{6}$$

provides the upper bound for the solution of problem (1) where such a bound is valid for all distributions defined by (3) and (4). Similarly, the solution of problem

$$(7) \qquad \min_{x\in X} \max_{H(\zeta_v|\zeta_{c(v)})\in G_{v,c(v)}} \int_V f^0(x,\omega_V) \prod_{v\in V} \mathrm{d}H(\omega_v|\omega_{c(v)})$$

provides the upper bound for the solution of problem (2).

The respective lower bounds are defined similarly by substituting max for min in (6) and (7).
In the next sections we consider the case when controlled parameters are fixed. Under this assumption problem (7) becomes:

$$\max_{H(\zeta_v|\zeta_{c(v)})\in G_{v,c(v)}} \int_\Omega f^0(\omega_V) \prod_{v\in V} \mathrm{d}H(\omega_v|\omega_{c(v)})$$

The rest of the paper is organized as follows: in Section 2 we give a bief introducton about Bayesian Networks and define a new class named Robust Bayesian Networks and introduce estimation problems on Robust Bayesian Networks. In Section 3 we describe two different approaches for solving estimation problems defined on Robust Bayesian Nets. Section 4 is devoted to the description of estimation algorithms based upon the approaches introduced in Section 3. In Section 5 we present a case study related to Integrated Circuits Manufacturing (ICM) together with numerical experiments. Finally in Section 6 is described a possible direction for further research.

2. Robust Bayesian Nets. In this section we gather basic definitions which will be used in the sequel. Starting from the definition of Bayesian Net we define Robust Bayesian Net (RBN). Suppose that $G = (V, E)$ is an acyclic directed graph with V being the set of nodes and $E \subseteq V \times V$ being the set of directed arcs. For each node $v \in V$ let us define respectively the set of parents and the set of descendants:

$$c(v) = \{w|(w,v) \in E\} \qquad d(v) = \{w|(v,w) \in E\}$$

Let us consider a set of random variables $\zeta_V = \{\zeta_v, v \in V\}$ indexed by nodes V and defined on appropriate probability space $(\Omega, \mathbf{B}, \mathbf{P})$. Suppose that W is an arbitrary subset of V and ζ_W is a vector which components are indexed by elements of W. The graph G is associated with ζ_V and the node v is associated with variable ζ_v of ζ_V in the sense that G describes probabilistic dependencies between different elements of ζ_V. More precisely, suppose that W is an arbitrary subset of nodes such that

$$W \subseteq V \setminus \{d(v) \cup v\}$$

Then ζ_v and ζ_W are conditionally independent given $\zeta_{c(v)}$, i.e.

$$\mathbf{P}(\zeta_v|\zeta_{W\cup c(v)}) = \mathbf{P}(\zeta_v|\zeta_{c(v)}). \tag{8}$$

DEFINITION 1 *A pair (ζ_V, G) which satisfies (8) is called a Bayesian Net.*

Suppose now that $H^*(\zeta_V)$ is the joint distribution of ζ_V. It follows from Definition 1 that

$$H^*(\zeta_V) = \prod_{v\in V} H^*(\zeta_v|\zeta_{c(v)}) \tag{9}$$

where $H^*(\zeta_v|\zeta_{c(v)})$ is the conditional distribution function of ζ_v given $\zeta_{c(v)}$. Now we are going to introduce the definition of *Robust Bayesian Net.*

Suppose that $\mathcal{H}_V$ is the set of all distributions representable in the form (9), i.e.

$$\mathcal{H}_V = \left\{ H(\zeta_V) \geq 0, \int_\Omega \mathrm{d}H(\omega_V) = 1, H(\zeta_V) = \prod_{v\in V} H(\zeta_v|\zeta_{c(v)}) \right\} \tag{10}$$

and $\mathcal{A}$ is some subset of $\mathcal{H}_V$. This subset contains all available information about the distribution $H^*(\zeta_V)$ of Bayesian net (ζ_V, G).

DEFINITION 2 *A triple $(\zeta_V, G, \mathcal{A})$ where ζ_V, G are the same as in Definition 1 and $\mathcal{A} \subseteq \mathcal{H}_V$ from (10) is called Robust Bayesian Net.*

In order to be useful this definition should be supplemented with meaningful examples of the distribution sets $\mathcal{A}$. One such example is constituted by *generalized moment constraints* (Betró *et. al* (1994)).

Suppose that we have the set of functions $f^k(\omega_V)$, $k = 1, 2, \ldots n$, then we can define the set $\mathcal{A}_1$ as follows:

$$\mathcal{A}_1 = \left\{ H | H \in \mathcal{H}_V, \int_\Omega f^k(\omega_V) \prod_{v\in V} \mathrm{d}H(\omega_v|\omega_{c(v)}) \leq 0, \; k = 1, 2, \ldots n \right\}. \tag{11}$$

Let us now define different estimation problems associated with Robust Bayesian Nets.

PROBLEM 1 *For a given Robust Bayesian net $(\zeta_V, G, \mathcal{A})$ we define the a priori estimation problem as that of finding the upper and lower bounds on the values of $\mathbf{E}f^0(\zeta_V)$ where $f^0(\zeta_V)$ is a known function of random vector ζ_V. This is equivalent to the solution of the following problems:*

$$\max_{H\in\mathcal{A}} \mathbf{E}f^0(\zeta_V) = \max_{H\in\mathcal{A}} \int_\Omega f^0(\omega_V) \prod_{v\in V} dH(\omega_v|\omega_{c(v)}) \tag{12}$$

$$\min_{H\in\mathcal{A}} \mathbf{E}f^0(\zeta_V) = \min_{H\in\mathcal{A}} \int_\Omega f^0(\omega_V) \prod_{v\in V} dH(\omega_v|\omega_{c(v)}) \tag{13}$$

Similarly we can define other important problems, e.g. aposteriori estimation and optimization problems. We formulate here the a posteriori estimation problem for the case of discrete random variables. The same formulation is valid for continuous random variables which have densities. In such a case we obtain the formulation of the problem by substituing densities for distributions.

PROBLEM 2 *For a given Robust Bayesian net* $(\zeta_V, G, \mathcal{A})$ *let* $\zeta_\mathcal{E} = \omega^*_\mathcal{E}$ *be the available statistical information and assume the case when random variables* ζ_V *are discrete. In such a case we have the following:*

$$H(\zeta_{V\setminus\mathcal{E}}|\zeta_\mathcal{E} = \omega^*_\mathcal{E}) = \frac{H(\zeta_{V\setminus\mathcal{E}}, \zeta_\mathcal{E} = \omega^*_\mathcal{E})}{H(\zeta_\mathcal{E} = \omega^*_\mathcal{E})} = \frac{H(\zeta_\mathcal{E} = \omega^*_\mathcal{E}|\zeta_{V\setminus\mathcal{E}})\, H(\zeta_{V\setminus\mathcal{E}})}{H(\zeta_\mathcal{E} = \omega^*_\mathcal{E})}$$

$$= \frac{1}{C} \prod_{v\in\mathcal{E}} H(\zeta_v = \omega^*_v|\zeta_{c(v)}) \prod_{w\in V\setminus\mathcal{E}} H(\zeta_w|\zeta_{c(w)}).$$

Then we define a posteriori estimation problem as that of finding the upper and lower bounds on the values of $\mathbf{E}f^0(\zeta_V)$ *where* $f^0(\zeta_V)$ *is a known function of random vector* ζ_V. *This is equivalent to the solution of the following problems:*

$$\max_{H\in A} \mathbf{E}\, f(\zeta_{V\setminus\mathcal{E}}|\omega^*_\mathcal{E}) =$$

$$= \max_{H\in A} \frac{1}{C} \int_\Omega f(\omega_V) \prod_{v\in\mathcal{E}} H(\omega^*_v|\omega_{c(v)}) \prod_{w\in V\setminus\mathcal{E}} dH(\omega_w|\omega_{c(w)}) \tag{14}$$

$$\min_{H\in A} \mathbf{E}\, f(\zeta_{V\setminus\mathcal{E}}|\omega^*_\mathcal{E}) =$$

$$= \min_{H\in A} \frac{1}{C} \int_\Omega f(\omega_V) \prod_{v\in\mathcal{E}} H(\omega^*_v|\omega_{c(v)}) \prod_{w\in V\setminus\mathcal{E}} dH(\omega_w|\omega_{c(w)}) \tag{15}$$

3. Lagrange relaxation and Ω-discretization. In this section we propose two different appoaches for solving problems (12) (13) (14) and (15). Let us now focus on problem (12). We propose two different approaches for analyzing the above estimation problem named: Lagrange relaxation and probabilities discretization.

The first approach relies on *Lagrange relaxation* and can be introduced as follows. Let $F^k(H) = \int_\Omega f^k(\omega) dH(\omega)$ $k = 1, 2, \ldots n$ and let $\mathcal{F}'$ be defined as follows:

$$\mathcal{F}' = \left\{F = (\mathcal{F}^0(H(\zeta_V)),\ F^1(H(\zeta_V)),\ \ldots,\ F^n(H(\zeta_V))) :\ H \in \mathcal{H}_V\right\}$$

then for general constraints (5) it is possible to show that $\mathcal{F}'$ is non-convex. Let us now consider the case when each constraint is related to a single random variable. In such a case we provide the following theorem.

THEOREM 1 *(Stella (1995)) Assume that* $f^k(\omega)\ \forall k = 0,1,\ldots,n$ *depends on a single variable* ζ_k *then the set* $\mathcal{F}'$ *is convex and* $\mathcal{F}' = Co\mathcal{Z}$ *where:* $\mathcal{Z} = \{z : z = (f^0(\omega), f'(\omega), \ldots f^n(\omega)),\ \omega \in \Omega\}$

PROOF: Let $F^k(H_1)$, $F^k(H_2)$ be any pair of points in $\mathcal{F}'$. Let us consider the following linear combination.

$$\lambda F^k(H_1)+(1-\lambda)F^k(H_2) = \int_{\Omega_v} f^k(\omega)(\lambda h^1(\zeta_v \mid \zeta_{c(v)})+(1-\lambda)h^2(\zeta_v \mid \zeta_{c(v)}))d\zeta_v.$$

We obtain a new distribution function H^* defined as follows:

$$H^* = (\lambda h^1(\zeta_v \mid \zeta_{c(v)}) + (1-\lambda)h^2(\zeta_v \mid \zeta_{c(v)})) \prod_{\omega \in V, \omega \neq v} h(\zeta_\omega \mid \zeta_{c(\omega)})$$

which generates $F^k(H^*) = \lambda F^k(H_1) + (1-\lambda)F^k(H_2)$. This implies that: $\mathcal{Z} \subseteq \mathcal{F}'$ $\mathcal{F}' \subseteq Co\mathcal{Z}$ $\mathcal{F}'$ is convex and then:

$$\mathcal{F}' = Co\mathcal{Z}. \tag{16}$$

This completes the proof. □

Using the results from Theorem 1 it is possible to show that problem (12) is equivalent to a linear programming problem obtained throught Lagrange relaxation as stated in the following theorem.

THEOREM 2 *(Stella (1995)) Assume that* Ω *is a compact set,* $f^k(\omega)$, $k = 0,1,\ldots,n$ *are continuous functions depending on a single variable and* $0 \in$ *int CoZ. Consider the following min-max problem:*

$$\min_{\lambda \in \Lambda^+} \max_{\omega \in \Omega} \left[f^0(\omega) - \sum_{k=1}^{n} \lambda_k f^k(\omega) \right] \tag{17}$$

where $\Lambda^+ = \{\lambda = (\lambda_0, \lambda_1, ..., \lambda_n),\ \lambda_k \geq 0\}$. *Let* $L(\lambda)$ *be defined as follows:*

$$L(\lambda) = \max_{\omega \in \Omega} \left[f^0(\omega) - \sum_{k=1}^{n} \lambda_k f^k(\omega) \right]$$

and

$$\Lambda^* = \{\lambda^* \in \Lambda^+ : L(\lambda^*) = \min_{\lambda \in \Lambda^+} L(\lambda)\}$$

$$\Omega^* = \left\{ \omega^* \in \Omega : f^0(\omega^*) - \sum_{k=1}^{n} \lambda_k f^k(\omega^*) = \max_{\omega \in \Omega} \left[f^0(\omega) - \sum_{k=1}^{n} \lambda_k f^k(\omega) \right] \right\}$$

Supp H^* *is the support set of distribution* H. *Then the solutions of both problem (12), (17) exist, and the optimal values of both problems are equal. Furthermore for each solution* H^* *for (12) there exists* $\lambda^* \in \Lambda^*$ *such that:*

$$supp\ H^* \subseteq \Omega^*(\lambda^*)$$

For the proof of Theorem 2 the reader is referred to Stella (1995).

Let us now present the Theorem upon which is based the second approach for solving problem (12) named probabilities discretization.

THEOREM 3 *(Stella (1995)) Consider problem (12) and assume that the functions* $f^k(\omega), k = 0, 1, \ldots, n$ *are* $\mathbf{B}$ *– measurable, than there exists a distribution function* $H(\zeta_V)$ *solution for the aforementioned problem such that* $H(\zeta_V) = \prod_{v \in V} H(\zeta_v|\zeta_{c(v)})$ *and such that every conditional distribution function* $H(\zeta_v|\zeta_{c(v)})$ *is discrete consisting of a support of at most* $n+1$ *distinct values.*

For the sake of brevity in this paper we do not report the proof for Theorem 3 which is mainly based on results coming from Kempermann (1968).

4. Estimation algorithms. In this section we present three stochastic optimization algorithms based on Theorems 2 and 3 respectively. Let us now present a numerical algorithm for solving problem (12) based on Theorem 2.

ALGORITHM 1 Step 1. Let s be the current iteration number and set $s = 0$. Furthermore let $l \in N$ be a generic number and let ϵ be any given suiably small quantity. Choose randomly a starting pair of points $\omega^{(s)} \in \Omega$ and $\lambda^{(s)} \in \Lambda$. Set $\overline{\omega}^s = \omega^s$ where $\overline{\omega}^s$ represents the optimal point at iteration s and $\varphi(\omega, \lambda) = f^0(\omega) - \sum_{k=1}^n \lambda_k f^k(\omega)$.

Step 2. Select randomly a point $\omega^{(s+1)} \in \mathcal{N}(\omega^{(s)})$ where $\mathcal{N}(\omega^{(s)})$ is some neighborhood of $\omega^{(s)}$.

Step 3. Select a new point $\lambda^{(s+1)}$ according to the following

$$\lambda_k^{s+1} = \max\left\{0; \lambda_k^s - \rho_k^s \frac{\partial \varphi(\omega : \lambda)}{\partial \lambda_k} \mid \omega = \overline{\omega}^s\right\} \quad \forall k = 1, 2, \ldots n$$

where ρ_k^s is the step at iteration s.

Step 4. Compute the following quantities

$$\varphi^{s+1} = \varphi(\omega^{s+1}; \lambda^{s+1}) = \varphi(\overline{\omega}^{s+1} : \lambda^{s+1}).$$

Step 5. If $\varphi^{(s+1)} \geq \overline{\varphi}^{(s+1)}$ then set $\overline{\varphi} = \varphi^{s+1}$ and $\overline{\omega}^{s+1} = \omega^{s+1}$ otherwise set $\overline{\varphi} = \overline{\varphi}^{s+1}$ and $\overline{\omega}^{s+1} = \overline{\omega}^s$.

Step 6. Compute the following quantity:

$$\gamma_k^{s+1} = \sum_{j=s+l-l}^{s+1} -f^k(\overline{\omega}^j).$$

If for each $\lambda_k^{s+1} > 0$ we have that $|\gamma_K^{s+1}| < \epsilon$ and for each $\lambda_k^{s+1} = 0$ we have that $\gamma_k^{s+1} > -\epsilon$ then the algorithm terminates.

A similar algorithm has been developed for the a posteriori estimation problem (14) based on Theorem 2.

ALGORITHM 2 *Step 1.* Let s be the current iteration number and set it equal to zero: $s = 0$. Furthermore let $l \in N$ be a generic number and let ϵ be any given suitably small quantity. Choose randomly a starting triplet of points $\omega^{(s)} \in \Omega$. $\Lambda^{(s)} \in \Lambda$ and $\mu^s \in M$. Set $\overline{\omega}^s = \omega^s$ where $\overline{\omega}^s$ represents the optimal point at iteration s and $\varphi(\omega, \lambda, \mu) = f^0(\omega) - \sum_{k=1}^{n} \lambda_k f^k(\omega) - \sum_{k=1}^{m} \mu_k g^k(\omega)$.

Step 2. Select randomly a point $\omega^{s+1} \in \mathcal{N}(\omega^{(s)})$ where $\mathcal{N}(\omega^{(s)})$ is some neighborhood of $\omega^{(s)}$.

Step 3. Select a new pair of point $\lambda^{(s+1)}$ and μ^{s+1} according to the following

$$\lambda_k^{s+1} = \max\left\{0, \lambda_k^s - \rho_k^s \frac{\partial \varphi(\omega : \lambda : \mu)}{\partial \lambda_k} \mid \omega = \overline{\omega}^s\right\} \quad \forall k = 1, 2, \ldots n$$

$$\mu_k^{s+1} = \max\left\{0, \mu_k^s - r_k^s \frac{\partial \varphi(\omega : \lambda : \mu)}{\partial \mu_k} \mid \omega = \overline{\omega}^s\right\} \quad \forall k = 1, 2, \ldots m$$

Step 4. Compute the following quantities

$$\varphi^{s+1} = \varphi(\omega^{s+1} : \lambda^{s+1} : \mu^{s+1}) \quad \overline{\varphi}^{s+1} = \varphi(\overline{\omega}^{s+1} : \lambda^{s+1} : \mu^{s+1}).$$

Step 5. If $\varphi^{s+1} \geq \overline{\varphi}^{(s+1)}$ then set $\overline{\varphi} = \varphi^{s+1}$ and $\overline{\omega}^{s+1} = \omega^{s+1}$ otherwise set $\overline{\varphi} = \overline{\varphi}^{s+1}$ and $\overline{\omega}^{s+1} = \overline{\omega}^s$.

Step 6. Compute the following quantities:

$$\gamma_k^{s+1} = \sum_{j=s+l-l}^{s+1} -f^k(\overline{\omega}^j) \qquad \delta_k^{s+1} = \sum_{j=s+l-l}^{s+1} -g^k(\overline{\omega}^j).$$

If for each $\lambda_k^{s+1} > 0$ we have that $|\gamma_k^{s+1}| < \epsilon$, for each $\lambda_k^{s+1} = 0$ we have that $\gamma_k^{s+1} > -\epsilon$, for each $\mu_k^{s+1} > 0$ we have that $|\delta_k^{s+1}| < \epsilon$ and for each $\mu_k^{s+1} = 0$ we have that $\delta_k^{s+1} > -\epsilon$ then the algorithm terminates.

The proof of convergence for Algorithms 1 and 2, which is not reported in this paper for the sake of brevity, is mainly based on results coming from Gaivoronski (1986) and sufficient conditions for the convergence of the algorithms 1 and 2 are:

$$(18) \qquad \rho_k^s \geq 0 \quad \sum_{s=1}^{+\infty} \rho_k^s = +\infty \quad \sum_{s=1}^{+\infty} (\rho_k^s)^2 < +\infty \quad \forall k = 1, 2, \ldots, n$$

$$(19) \qquad r_k^s \geq 0 \quad \sum_{s=1}^{+\infty} r_k^s = +\infty \quad \sum_{s=1}^{+\infty} (r_k^s)^2 < +\infty \quad \forall k = 1, 2, \ldots, n$$

For further details the reader is refered to Stella (1995).

Finally let us present a numerical algorithm for solving problem (12) based on Theorem 3.

ALGORITHM 3 Step 1. For each root node $i \in V$ (i.e. $c(i) = \emptyset$) randomly select $n+2$ different points $\omega_i^{(j)}$, $j = 1, 2, ..., n+2$. For each one of the remaining nodes $i \in V$ and for each configuration of its parent nodes $\omega_{c(i)}$ randomly select $n+2$ different points $\omega_i^{(j)}$, $j = 1, 2, ..., n+2$. Set the number of current iteration $h = 0$.

Step 2. For each node $i \in V$ and for each configuration of its parent nodes $\omega_{c(i)}$ set:

$$p_{j_{i,c(i)}} = \frac{1}{(n+2)} \qquad j = 1, 2, ..., n+2.$$

Step 3. Select a node $i \in V$ not yet selected during the current cycle and increase the current number of selected nodes K, $K = K + 1$. Solve the following linear programming problem:

$$\max_{p_{j_{i,c(i)}}, \sigma} \sum_{j_{i,c(i)}=1, i \in V}^{n+2} p_{j_{i,c(i)}} f^0(\omega^I) \prod_{l \in V, l \neq i} p_{j_{l,c(l)}} - M\sigma$$

subject to:

$$\sum_{j_{i,c(i)}=1, i \in V}^{n+2} p_{j_{i,c(i)}} f^k(\omega^I) \prod_{l \in V, l \neq i} p_{j_{l,c(l)}} \leq \sigma \qquad \forall k = 1, 2, \ldots, n$$

$$\sum_{j_{i,c(i)}=1}^{n+2} p_{j_{i,c(i)}} = 1 \qquad \forall \omega_{c(i)}$$

where M is some given large number. If the auxiliary variable $\sigma = 0$ then go to step 4, else if $K < |V|$ go to step 3, else set K=0 and go to step 3.

Step 4. Update the number of current random searchs, $h = h + 1$. For each node $i \in V$ and for each configuration of its parent nodes $\omega_{c(i)}$ select $\omega_j^{(i)}$ such that $p_{j_{i,c(i)}} = 0$. Then replace $\omega_j^{(i)}$ with $\omega_j^{(i)*} \in \mathcal{N}(\omega_j^{(s)})$ where $\mathcal{N}(\omega_j^{(s)})$ is some neighborhood of $\omega_j^{(i)}$.
Solve the following linear programming problem:

$$\max_{p_{j_{i,c(i)}}} \sum_{j_{i,c(i)}=1, i \in V}^{n+2} p_{j_{i,c(i)}} f^0(\omega^I) \prod_{l \in V, l \neq i} p_{j_{l,c(l)}}$$

subject to:

$$\sum_{j_{i,c(i)}=1, i\in V}^{n+2} p_{j_{i,c(i)}} f^k(\omega^I) \prod_{l\in V, l\neq i} p_{j_{l,c(l)}} \leq 0 \quad \forall k = 1,2,\ldots,n$$

$$\sum_{j_{i,c(i)}=1}^{n+2} p_{j_{i,c(i)}} = 1 \quad \forall \omega_{c(i)}$$

Step 5. Update the sample space for the node providing the best objective function improvement.

Step 6. If $h < N$ go to step 4, otherwise the algorithm terminates.

5. Integrated circuit robustness analysis. In this section we provide a description of the robustness analysis of Integrated Circuits (ICs) based on the theory and algorithms described in preceding sections. We first provide a brief description of Integrated Circuit Manufacturing (ICM) and then report a numerical experiment related to Algorithm 1.

The silicon wafer production process is characterized by a high number of steps (about 80-120, according to the microcircuit to be produced). This makes it very difficult for process engineers to diagnose the causes of failure. On each silicon wafter (Figure 1) are located several chips (about 300 up to 400) and five test structures hereafter referred as "test patterns". These structures contain a variety of elements which can be contacted via probe pads, e.g. Resietor, Contact, Metal Comb, Diode, Bipolar, Transistor and MOS Transistor. In order to control the IC fabrication process test data for both electrical and non-electrical parameters are collected during and after processing. These measurements include:

In-process (IP) measurements: taken at several steps during the process. They are normally utilized to control a specific process step. Three major type of measures are carried out: sheet resistance of layers Rs. thickness of grown or deposited layed d and critical dimension CD the line width of photolitographic structure.

Parametric Control Monitors (PCM) tests: performed immediately after the whole process is completed. The measurements performed on test patterns are compared to previously established control limits in order to draw conclusions regarding the succesful completion of the processing cycle.

Wafer Final Test (WFT): usually performed in two steps. In the first step named "wafer test" ICs on unsawn wafers are tested in order to determine whether they satisfy the IC specification limits. Based upon the test results the ICs are classified into groups having different quality levels and packaged. Then, the second stage of IC testing is performed on packaged microcircuits where similar or identical tests as in the first stage are performed. Notice

that the WFT test procedure allows also the detection of defectiveness while PCM does not.

Based upon the ICM case study we performed a wide set of numerical experiments related to Algorithms 1, 2 and 3. More precisely we analyzed the behavior of the above mentioned algorithms w.r.t. some parameters of interest such as algorithm step and objective function type. For brevity we only describe some results related to Algorithm 1. More information about numerical experiments can be found in Stella (1995).

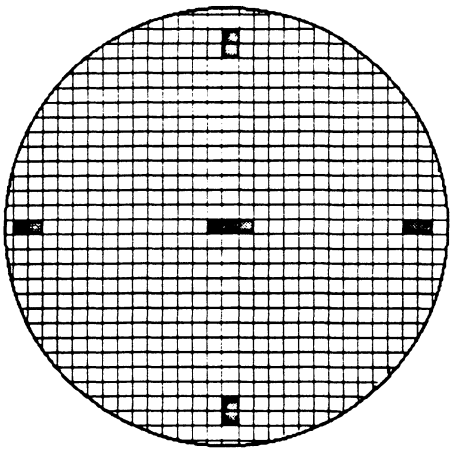

Figure 1: *Silicon Wafer*

EXAMPLE. The case study we are going to present in the rest of the paper is related to the DMOS microcircuit. The Bayesian net model related to the DMOS microcircuit (Figure 2), has been obtained by combining expert knowledge together with statistical estimation procedures. In Table 1

Node	Description	Node	Description
v_1	Epi sheet resistance	v_{10}	Hfe NPN P-body base
v_2	P-body sheet resistance	v_{11}	LDMOS Vt
v_3	N+ As sheet resistance	v_{12}	LDMOS Vt
v_4	Gate oxide on epi capacit.	v_{13}	Bvcbo HV Power NPN-PBB
v_5	Si-poly strip width	v_{14}	BVcbo NPN P-body base
v_6	Drain estension sheet res.	v_{15}	Low Leakage diode BV
v_7	Vt shift due to B eff.	v_{16}	Power LDMOS Ron
v_8	Vbe NPN P-body base	v_{17}	BVceo NPN P-body base
v_9	BVebo NPN P-body base	v_{18}	Power VDMOS Ron

Table 1: *Node-Parameter Correspondency Table*

we report the physical meaning for each node described in Figure 2.

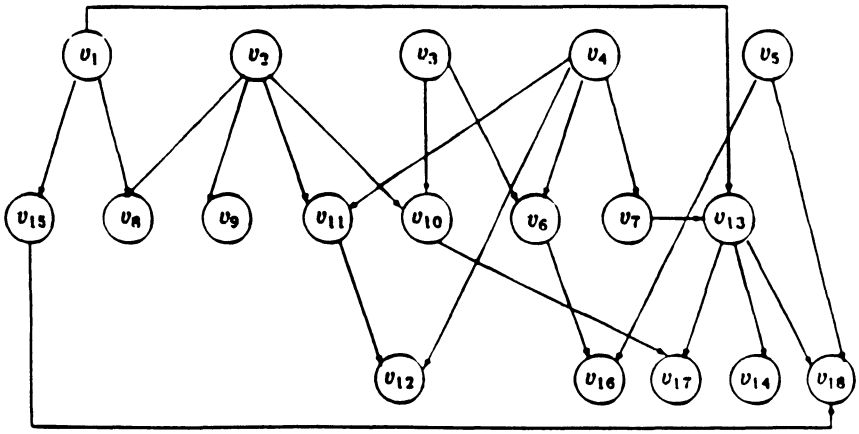

Figure 2: *Robust Bayesian Net for ICM-DMOS Robustness Analysis*

It is worth observing that the case study we are going to present is specifically related to PCM parameters but in principle there is not reason why a more general model involving IP and WFT parameters cannot be analyzed using our approach.

In order to better clarify the nature of our problem let us specify that each random variable can take values within the range [-1.1] according to a gaussian distirbution. Furthermore in order to define probability constraints we split the range [-1.1] as follows:

$$I_1 = [-1.0) \quad I_2 = [0.1].$$

According to the above range splitting and for any given node v_j we consider generalized moment constraints of the form:

$$\mu_j - \epsilon_\beta \frac{\sigma_j}{\sqrt{M}} \leq P(\zeta_j \in I_i \mid \zeta_{c(j)} \in I_{c(j)}) \leq \mu_j + \epsilon_\beta \frac{\sigma_j}{\sqrt{M}}. \tag{20}$$

Finally we present the case when the objective function is defined on node $v_{13}(P(\zeta_{13} \in I_2))$. In order to describe and discuss the behavior of algorithm 1 we have to introduce some new quantities. More precisely let S be the algorithm iteration number. A, B, C be real variables and ρ_s be the step of algorithm 1. We consider the case when the algorithm step is defined as $\rho_s = 1/(SA + B)^C$, where $A = 0.3, B = 50$ and $C = 1$.

In Figure 3 the behavior of Algorithm 1 is reported int he case when we have to estimate a lower bound for our objective function.

The numerical experiment was performed assuming that in (20) the σ_j is set to zero. This enables us to compute the correct solution to our problem

which in Figure 3 is shown as means of a straight line. It can be observed that our algorithm converges quite quickly to this solution and around iteration 600 the precision is quite appreciable. We also performed numerical experiments assuming $\sigma_j \neq 0$ but in such a case it is not simple to evaluate the algorithm behavior.

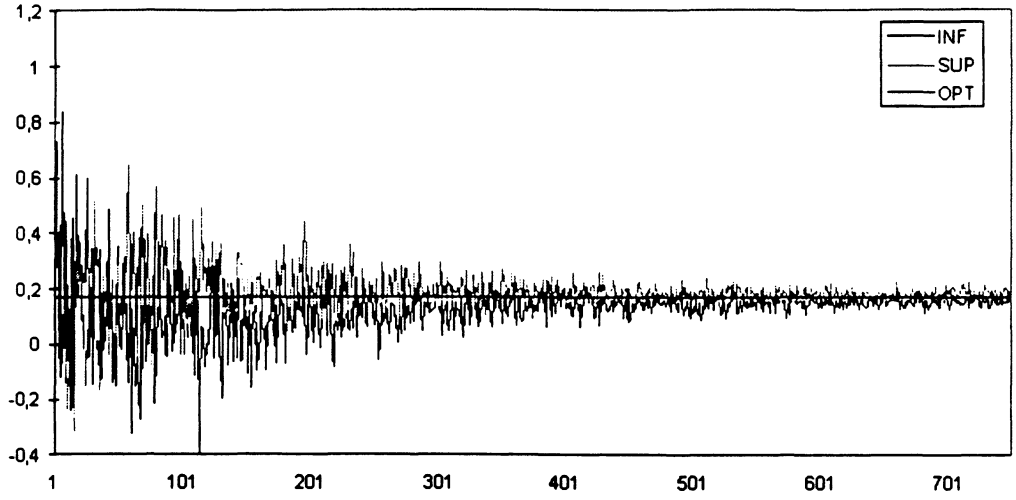

Figure 3: *Lagrange Multipliers Algorithm behavior*

6. Direction for further research. In this paper we have considered some optimization and estimation problems on Robust Bayesian nets, i.e. nets in which we have uncertainty about conditional distributions of random variables associated with vertices of acyclic directed graphs. We assumed that the graph which defines the strucure of Robust Bayesian nets is known. However, in many applications it is not completely true and we have also uncertianty about the strucutre (Lauritzen *et al.* (1993)). This is a very important problem which is one of the directions for further research. We can give here only very brief outlines of our approach.

Let us consider, for example, an extension for this case of the a priori estimation problem (1). Suppose that we know the set $\mathcal{G}$ of possible structures of conditional dependencies and for each $G \in \mathcal{G}$ we have information $\mathcal{A}(G)$ about conditional distributions. In this case we can study robustness properties also with respect to structural uncertianty as follows:

PROBLEM 1. *For a given set of possible structures $\mathcal{G}$ we define the a priori esimation problem with structural uncertainty as the problem of finding the structures $\underline{G} \in \mathcal{G}$ and $\overline{G} \in \mathcal{G}$ corresponding to Robust Bayesian nets $(\zeta_V, \underline{G}, \mathcal{A}(\underline{G}))$, $(\xi v, G, \mathcal{A}(\underline{G}))$ and the upper and lower bounds of the values of $\mathbf{E} f^0(\zeta_V)$ from the solution of the following problems:*

$$(21)\quad \max_{G\in\mathcal{G}} \max_{H\in\mathcal{A}(\mathcal{G})} Ef^0(\zeta_V) = \max_{G\in\mathcal{G}} \max_{H\in\mathcal{A}(G)} \int_\Omega f^0(\omega_V) \prod_{v\in V} dH(\omega_v \mid \omega_{c(v)})$$

$$(22)\quad \min_{G\in\mathcal{G}} \min_{H\in\mathcal{A}(G)} Ef^0(\zeta_V) = \min_{G\in\mathcal{G}} \min_{H\in\mathcal{A}(G)} \int_\Omega f^0(\omega_V) \prod_{v\in V} dH(\omega_v \mid \omega_{c(v)})$$

This problem is more difficult numerically than the problems (12), (13) and algorithms which exploits its structure should be developed for its solution.

Acknowledgements. The authors are grateful to the anonymous referees for useful comments and suggestions which have contributed to the improvement of the paper.

REFERENCES

BETRÒ, B. and GUGLIELMI, A. (1994). An algorithm for robust Bayesian analysis under generalized moment conditions. Technical Report IAMI 94.6, CNR-IAMI, Milano.

DALL'AGLIO, M. and SALINETTI, G. (1994). Bayesian robustness, moment problem and semi-finite linear programming. *Manuscript.*

ERMOLIEV, Yu.M., GAIVORONSKI, A. and NEDEVA, C. (1985). Stochastic optimization problems with incomplete information on distribution functions. *SIAM J. Control Optim.*, **23**, 697-716.

GAIVORONSKI, A. (1986). Linerization methods for optimization for optimization of functionals which depend on probability measures. *Math. Progr. Study.* **28**, 157–181.

GAIVORONSKI, A. and STELLA, F. (1994). A class of stochastic optimization algorithms applied to some problems in Bayesian statistics. *Operations Research Proceedings 1994*, 65-69.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intell.*, **6**, 721-741.

KEMPERMANN, J.H.B. (1968). The general moment problem, a geometric approach. *Ann. Math. Statist.*, **23**, 93-122.

LAURITZEN, S.L., SPIEGELHALTER, D.J., DAWID A.P. and COWELL, R.G. (1993). Bayesian analysis in expert systems. *Statistical Science*, **8**, 219-247.

PEARL, J. (1988). Probabilistic reasoning in intelligent systems. *Morgan Kaufmann*, San Mateo, CA.

STELLA, F. (1995). Problemi di ottimizzazione su reti Bayesiane. *Ph.D. Thesis in Computational Mathematics and Operations Research*, Dept. of Computer Sciences, Universitá Statale degli Studi di Milano, Milano.

ITALTEL
CASTELLETTO DI SETTIMO MILANESE
I-20129 MILANO, ITALY

DIPARTIMENTO DI SCIENZE
DELL'INFORMAZIONE
UNIVERSITÁ DEGLI STUDI
DI MILANO
VIA COMELICO, 39-41
I-20129 MILANO, ITALY