Sensitivity Analysis of Markovian Models

Theodore Charitos and **Linda C. van der Gaag**
Department of Information and Computing Sciences, Utrecht University
P.O Box 80.089, 3508 TB Utrecht, The Netherlands
{theodore,linda}@cs.uu.nl

Abstract

Sensitivity analysis of Markovian models amounts to computing the constants in polynomial functions of a parameter under study. To handle the computational complexity involved, we propose a method for approximate sensitivity analysis of such models. We show that theoretical properties allow us to reason for the present time using just few observations from the past with small loss in accuracy. The computational requirements of our method render sensitivity analysis practicable even for complex Markovian models. We illustrate our method by means of a sensitivity analysis of a real-life Markovian model in the field of infectious diseases.

Introduction

Whether estimated from data or assessed by experts, the parameters of a Markovian model tend to be inaccurate to at least some degree, due to incompleteness of data and partial knowledge of the domain under study. These inaccuracies may affect the output of the model. The effects of inaccuracies in the parameters of a graphical model on its output, can be investigated by subjecting the model to a *sensitivity analysis* (Laskey, 1995; Coupé and Van der Gaag, 2002; Chan and Darwiche, 2002). For a Markovian model, performing such an analysis amounts to stepwise varying each parameter separately and studying the effects on the output probability of interest. Previous work on sensitivity properties of Markovian models with a single process has shown that the sensitivity functions are quotients of two functions that are polynomial in a parameter under study (Charitos and Van der Gaag, 2004). The order of these polynomial functions is linear in the time scope that is taken into consideration, and establishing these functions is highly demanding from a computational point of view. We now generalise these results to all types of Markovian model and present an approximate method for sensitivity analysis that reduces the runtime requirements involved yet incurs only a small loss in accuracy. Our method is based on theoretical properties of Markovian models and can lead to substantial time and space savings in the computations involved. In addition, we present a method for approximating the functional form of a sensitivity function to allow for further computations. The method is based upon least-squares approximation and is showing promising results in our experiments. We illustrate our methods by means of a sensitivity analysis of a real-life Markovian model in the field of infectious diseases.

Markovian models

Sequential statistical models for reasoning about stochastic processes include hidden Markov models (HMMs) and dynamic Bayesian networks (DBNs); in the sequel, we assume that these models are *Markovian* in the sense that the future state of the modelled process is assumed to be independent of the past state given its present state. An HMM is a statistical model $H = (X, Y, A, O, \Gamma)$ that can be looked upon as an extension of a finite homogeneous Markov chain, including observable variables that depend on the hidden variable. We use X_n to denote the hidden variable at time step n, with states x_i^n, $i = 1, \ldots, l$, $l \geq 2$; the *transition matrix* for X_n is denoted as $A = \{a_{i,j}\}$ with elements $a_{i,j} = p(X_{n+1} = x_j^{n+1} \mid X_n = x_i^n)$, $i, j = 1, \ldots, l$, for all n. We denote the observable variables by Y_n, with values y_j, $j = 1, \ldots, m, m \geq 2$, that are generated from the state of the hidden variable according to a time-invariant *observation matrix* $O = \{o_{i,j}\}$ with $o_{i,j} = p(Y_n = y_j \mid X_n = x_i^n)$, $i = 1, \ldots, l$, $j = 1, \ldots, m$, for all n. Finally, we denote by $\Gamma = \{\gamma_i\}$ the *initial probability vector* for the hidden variable, with $\gamma_i = p(X_1 = x_i^1)$, $i = 1, \ldots, l$. A DBN can be looked upon as an extension of an HMM, capturing a process that involves a collection of hidden variables. A DBN is a graphical model that encodes a joint probability distribution on a set of stochastic variables, explicitly capturing the temporal relationships between them. DBNs are usually assumed to be time invariant, which means that the topology and the parameters of the model per time step and across time steps do not change.

Applying a Markovian model usually amounts to computing marginal probability distributions for the hidden variables at different times. In this paper, we focus on *monitoring*, which is the task of computing these distributions for some time step n given the observations that are available up to and including that time step. For HMMs, the *forward-backward* algorithm is available for this task (Rabiner, 1989). For DBNs, Murphy (2002) introduced the interface algorithm as an extension of the junction-tree algorithm for probabilistic inference in graphical models in gen-

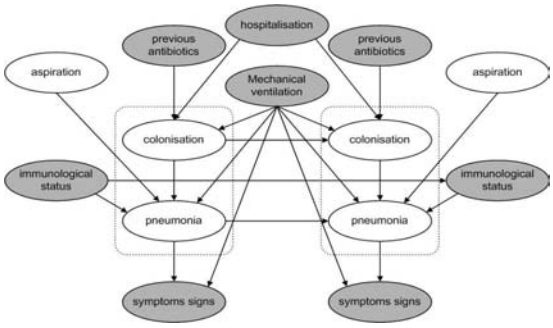


Figure 1: The dVAP model for the diagnosis of VAP for two consecutive time steps; clear nodes are hidden, shaded nodes are observable. The dashed boxes indicate the hidden processes of the model.

eral. The interface algorithm effi ciently exploits the concept of *forward interface*, which is the set of variables at time step n that affect some variables at time step $n+1$ directly; in the sequel, we use $\boldsymbol{FI}$ to denote this forward interface. Based on this concept, the interface algorithm requires two steps: a construction step and a numerical step. In the construction step, a junction tree is created for two consecutive time steps excluding all non-forward interface nodes and their incident arcs from the fi rst time step . In the numerical step, the clique that contains the forward interface serves as the root node for the computations. The computational complexity of the algorithm is exponential in the number of hidden variables and for large models can be prohibitive.

Throughout the paper we will use the *dVAP network* for illustration purposes. This network is a real-life Markovian model for diagnosing ventilator-associated pneumonia (VAP) in patients in an intensive care unit (ICU) and is destined for use in clinical practice (Charitos *et al., 2005*). The model has been developed and refi ned with the help of a single infectious disease specialist and has been evaluated for a period of 10 days on 20 patients from the ICU of the University Medical Centre Utrecht in the Netherlands; 5 of these patients were diagnosed with VAP. The network includes two interacting hidden processes (*colonisation* and *pneumonia*), three input processes (summarised in *immunological status*), three input observable variables (*hospitalisation*, *mechanical ventilation*, and *previous antibiotics*) and one hidden input variable (*aspiration*), and seven output observable variables (summarised in *symptoms-signs*). Per time step, representing a single day, the model includes 30 variables. Each of the interacting processes consists of seven subprocesses that are a-priori independent. The transition matrices of these processes are only moderately stochastic. Figure 1 shows the dVAP network in a compact way.

Sensitivity analysis revisited

Sensitivity analysis has been studied in the last decade in the context of Bayesian networks (BNs) (Laskey, 1995; Van der Gaag and Renooij, 2001; Coupé and Van der Gaag, 2002). It amounts to establishing, for each of the network's parameters, a function that expresses a given output probability of interest in terms of a parameter under study. We take the posterior probability $p(b \mid e)$ for our probability of interest, where b is a specifi c value of the variable B and e denotes the available evidence; we further let $\theta = p(h_i \mid \pi)$ be our parameter under study, where h_i is a value of the variable H and π is a specifi c combination of values for the parents of H. Sensitivity analysis now amounts to establishing the *sensitivity function* that describes $p(b \mid e)$ in terms of θ; we write $p(b \mid e)(\theta)$ for the function, thereby expressing the (algebraic) dependency of $p(b \mid e)$ upon θ. If we assume that the other parameters $p(h_j \mid \pi), h_j \neq h_i$, specifi ed for H are co-varied proportionally according to

$$p(h_j \mid \pi)(\theta) = \begin{cases} \theta & \text{if } j = i \\ p(h_j \mid \pi) \cdot \frac{1-\theta}{1-p(h_i \mid \pi)} & \text{otherwise} \end{cases}$$

for $p(h_i \mid \pi) < 1$, then the sensitivity function is a quotient of two linear functions in θ, that is,

$$p(b \mid e)(\theta) = \frac{p(b,e)(\theta)}{p(e)(\theta)} = \frac{c_1 \cdot \theta + c_0}{d_1 \cdot \theta + d_0}$$

where c_1, c_0, d_1 and d_0 are constants with respect to θ (Coupé and Van der Gaag, 2002). Under the assumption of proportional co-variation, therefore, any sensitivity function is characterised by at most three constants. Note that for parameters of which the probability of interest is algebraically independent, the sensitivity function simply equals the posterior probability $p(b \mid e)$; any computations can therefore be restricted to the *sensitivity set* for the variable of interest. The most effi cient scheme for sensitivity analysis to date (Kjaerulff and Van der Gaag, 2000) is based on the junction-tree algorithm. This scheme requires an inward propagation for processing evidence and a single outward propagation in the junction tree for establishing the constants of the sensitivity functions for all parameters per output probability. It builds on the idea that the expressions for $p(b,e)(\theta)$ and $p(e)(\theta)$ can be derived from the potential of a clique containing both the variable and the parents to which the parameter θ pertains.

Sensitivity properties of Markovian models

In a sensitivity analysis of a Markovian model, the probability of interest typically is the probability of a specifi c state of some hidden variable at time step $n > 1$. The parameter can be any parameter of the model, such as a transition probability or an observation probability. The main difference with sensitivity analysis of BNs is that a parameter occurs multiple times in a Markovian model. Previous work on sensitivity analysis of HMMs showed that the functions involved again are polynomials or quotients of polynomials, yet now of higher order (Charitos and Van der Gaag, 2004). In the sequel, we briefly review these results and generalise them to DBNs.

We begin by considering a Markovian model for which no evidence has been entered as yet. For an HMM, the probability of interest is the prior probability $p(x_r^n)$ of some state x_r of the hidden variable X_n. Let $\theta_a = a_{i,j} \in A$ be a transition parameter in the model. Then,

$$p(x_r^n)(\theta_a) = c_{n,r}^{n-1} \cdot \theta_a^{n-1} + \ldots + c_{n,r}^1 \cdot \theta_a + c_{n,r}^0$$

where $c^{n-1}_{n,r}, \ldots, c^0_{n,r}$ are constants with respect to θ_a dependent on time n. We thus have that the sensitivity function that expresses the prior probability $p(x^n_r)$ at time step n in terms of the transition parameter θ_a is a polynomial of order $n-1$ in this parameter. For an initial parameter $\theta_\gamma = \gamma_i \in \Gamma$, the function is linear :

$$p(x^n_r)(\theta_\gamma) = c^1_{n,r} \cdot \theta_\gamma + c^0_{n,r}$$

where $c^1_{n,r}$ and $c^0_{n,r}$ are constants with respect to θ_γ. Without any evidence, the probability of interest is algebraically independent of any observation probability.

We now assume that some evidence has been entered into the model; we use e_n to denote the combined evidence up to and including time step n. We consider again the probability of interest $p(x^n_r \mid e_n)$. Let $\theta_a = a_{i,j} \in A$ again be a transition parameter in the model. Then,

$$\frac{p(x^n_r, e_n)(\theta_a)}{p(e_n)(\theta_a)} = \frac{c^{n-1}_{n,r} \cdot \theta^{n-1}_a + \ldots + c^1_{n,r} \cdot \theta_a + c^0_{n,r}}{d^{n-1}_{n,r} \cdot \theta^{n-1}_a + \ldots + d^1_{n,r} \cdot \theta_a + d^0_{n,r}}$$

where $c^{n-1}_{n,r}, \ldots, c^0_{n,r}$, $d^{n-1}_{n,r}, \ldots, d^0_{n,r}$ are constants with respect to θ_a. We thus have that the sensitivity function that expresses the posterior probability $p(x^n_r \mid e_n)$ in terms of the transition parameter θ_a is a quotient of two polynomials in θ_a of order $n-1$. For an observation parameter $\theta_o = o_{i,j}$, the sensitivity function becomes

$$\frac{p(x^n_r, e_n)(\theta_o)}{p(e_n)(\theta_o)} = \frac{c^b_{n,r} \cdot \theta^b_o + \ldots + c^1_{n,r} \cdot \theta_o + c^0_{n,r}}{d^n_{n,r} \cdot \theta^n_o + \ldots + d^1_{n,r} \cdot \theta_o + d^0_{n,r}}$$

where $b = n$ if $r = i$ and $b = n-1$ otherwise; $c^b_{n,r}, \ldots, c^0_{n,r}$, $d^n_{n,r}, \ldots, d^0_{n,r}$ are constants with respect to the parameter θ_o. The order of the polynomials involved thus grows linearly with n. For an initial parameter θ_γ we have that the sensitivity function is a quotient of two linear functions in this parameter. For probabilities of interest belonging to any possible time step $n_o < n$ or $n_o > n$, similar results hold (Charitos and Van der Gaag, 2004).

The previous results are readily generalised to DBNs. Upon doing so, we will explicitly take into account the sensitivity set for the variable of interest B_n given the evidence e_n, denoted as $Sens(B_n, e_n)$. Note that the concept of sensitivity set was used implicitly for HMMs, where we argued for example that the sensitivity function for an observation parameter is a constant function as long as no evidence had been entered. In a DBN, we consider the posterior probability of interest $p(b^n_r \mid e_n)$ of the state b_r of the hidden variable B_n given the (possibly empty) evidence e_n. Then,

- for any variable $H_n \in Sens(B_n, e_n)$, the sensitivity function expressing $p(b^n_r \mid e_n)$ in $\theta = p(h^n_i \mid \pi)$ is a quotient of two polynomials of order $n-1$ if $H_n \in \boldsymbol{FI}$, or of order n otherwise;
- for any variable $H_n \notin Sens(B_n, e_n)$, the sensitivity function expressing $p(b^n_r \mid e_{n_o})$ in $\theta = p(h^{n_o}_i \mid \pi), n_o < n$, is a quotient of a polynomial of order $n - n_o$ in the numerator and a polynomial of order n_o in the denominator.

As an example, Figure 2 depicts the effect of varying the parameter $\theta = p(leucocytosis = yes \mid pneumonia = yes)$ on the probability distribution for pneumonia at day 10 given evidence e_{10} for a specifi c patient in the dVAP network. The depicted sensitivity function is a quotient of two polynomials of order 10 each.

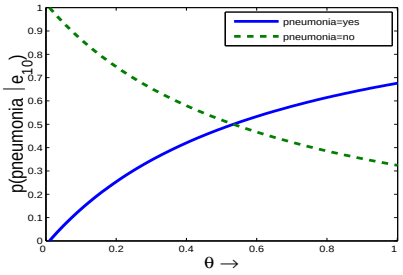


Figure 2: The sensitivity functions expressing the probabilities of pneumonia given e_{10} in terms of the parameter $\theta = p(leucocytosis = yes \,|\, pneumonia = yes)$.

To compute the constants in the sensitivity functions for a probability of interest in a DBN, we combine the interface algorithm with the scheme for sensitivity analysis from Kjaerulff and Van der Gaag (2000). Further details of this scheme are out of the scope of this paper.

Decreasing the computational requirements

The number of constants in the sensitivity functions of a DBN and the complexity of the propagations required to compute these constants grows linearly with n. For a large time scope, therefore, sensitivity analysis can become quite hard. We now propose to reduce the order of the polynomials and thereby the runtime requirements for their computation. We present an approximate technique for sensitivity analysis that builds on the concept of *mixing rate* of a Markov process. This concept has also been successfully used for approximate inference in large DBNs (Boyen, 2002). Informally speaking, when two different probability distributions are processed through a stochastic matrix, they become closer to one another. Based on this observation, we reduce the number of time steps for which perform inference upon computing the sensitivity functions.

Contraction of a single process

We consider two probability distributions μ and μ' over the same variable W. Conditioning on a set of observations is known to never increase the relative entropy of these distributions. Denoting the conditioning on a given set of observations by $o(\cdot)$, we thus have that

$$D[o(\mu) \| o(\mu')] \leq D[\mu \| \mu'] \qquad (1)$$

where D stands for the relative entropy. Now, consider the extreme case where μ and μ' have their probability mass on two different states w_i and w_k respectively. We denote by $A(\cdot)$ the distribution that results from processing through the transition matrix A. Even though μ and μ' do not agree on any state, processing through the transition matrix will cause them to place some mass on some state w_j. They then agree for a mass of $min\big[A(\mu(w_j; w_i)), A(\mu'(w_j; w_k))\big]$ on

that state w_j. Based on this property, the *minimal mixing rate* of the matrix A is defined as (Boyen, 2002):

$$\delta_A = \min_{i,k} \sum_j \min \left[A(\mu(w_j; w_i)), A(\mu'(w_j; w_k)) \right]$$

Given the minimal mixing rate of a transition matrix A, the following theorem now holds (Boyen, 2002):

$$D[A(\mu) \| A(\mu')] \leq (1 - \delta_A) \cdot D[\mu \| \mu']$$

We say that the stochastic process with transition matrix A *contracts* with probability δ_A. Combining equation (1) with the previous theorem we conclude that

$$D[A(o(\mu)) \| A(o(\mu'))] \leq (1 - \delta_A) \cdot D[\mu \| \mu']$$

Performing conditioning on two different distributions and transitioning them, will therefore result in two new distributions whose distance in terms of relative entropy is reduced by a factor smaller than one. Now, if we perform conditioning and transitioning on the resulting distributions and continue in this way, we are guaranteed that after some time steps there will be no longer any difference. The distance between the distributions in fact decreases exponentially with rate $(1 - \delta_A)$.

Our approximate method for sensitivity analysis now builds on the contraction property reviewed above. Suppose that we are interested in the probability of some state of the hidden variable X_n at time step n. After entering the available evidence e_n into the model, we can compute the exact posterior distribution $p(X_n \mid e_n)$. Building on the contraction property, however, we can also compute an approximate distribution $\widetilde{p}(X_n \mid e_n)$ starting from time step n_ϕ, with $1 < n_\phi < n$, without losing too much accuracy. We now define the *backward acceptable window* $\omega^\phi_{n,\epsilon}$ for time step n with a specified level of accuracy ϵ, to be the number of time steps we need to use from the past to compute the probability distribution of the hidden variable at time step n within an accuracy of ϵ. The following schematic figure illustrates our concept of the backward acceptable window:

$$\underbrace{\{1, \ldots, n_\phi, \ldots, n\}}_{\text{total time scope } n} \longrightarrow \underbrace{\{n_\phi, \ldots, n\}}_{\omega^\phi_{n,\epsilon}}$$

We now propose to perform sensitivity analysis for time step n considering only the backward acceptable window $\omega^\phi_{n,\epsilon}$. Note that the resulting functions then include polynomials of order $O(n - n_\phi)$ rather than of order $O(n)$ compared to the true functions.

For a given level of accuracy ϵ, we can determine the maximum value of n_ϕ for which

$$\begin{aligned} &D[p(X_n \,|\, e_n) \| \widetilde{p}(X_n \,|\, e_n)] \leq \\ &(1 - \delta_A)^{n - n_\phi} \cdot D[p(X_{n_\phi} \,|\, e_{n_\phi}) \| p(X_1)] \leq \epsilon \end{aligned}$$

where $\widetilde{p}(X_n \mid e_n)$ denotes the approximate distribution of X_n that is computed using $\omega^\phi_{n,\epsilon}$. Solving for n_ϕ, we find that

$$n_\phi \leq n - \left\lfloor \frac{\log \left(\epsilon / D[p(X_{n_\phi} \,|\, e_{n_\phi}) \| p(X_1)] \right)}{\log(1 - \delta_A)} \right\rfloor \qquad (2)$$

where $\lfloor \cdot \rfloor$ stands for the integer part. Starting from $n_\phi = n$ and decreasing the value of n_ϕ one step at a time, we can readily establish the value of n_ϕ that first satisfies equation (2). To this end, the interface algorithm needs to have computed and stored the exact posterior distributions $p(X_{n_o} \mid e_{n_o})$ for all $n_o \leq n$, given evidence e_{n_o}.

In view of sensitivity analysis, we observe that the value of n_ϕ that is established as outlined above, is based on the original values of all parameters of the model under study. We further observe that the minimal mixing rate δ_A used in the computation of n_ϕ is algebraically dependent only of the model's transition parameters. Using $\omega^\phi_{n,\epsilon}$ based upon n_ϕ for sensitivity analysis, therefore, is guaranteed to result in approximate sensitivity functions within accuracy of ϵ for any non-transition parameter. For transition parameters, this guarantee does not hold in general. We note, however, that for the original value of a transition parameter, the difference between the true probability of interest and the approximate one is certain to be smaller than ϵ. Since the value n_ϕ changes with δ_A in a stepwise manner only, this property holds for a range of values for the parameter. Our experimental results using the backward acceptable window with sensitivity analysis of the dVAP model in fact show that for all possible values of the transition parameters good approximations are found; we return to this observation presently.

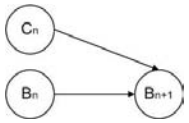
The procedure to compute the optimal value n_ϕ requires at most n computations of equation (2) and thus is not very demanding from a computational point of view. We recall, however, that for the computation of n_ϕ, the interface algorithm needs to have established the exact posterior distributions given the available evidence. Now in a full sensitivity analysis, the effects of parameter variation are being studied for a number of evidence profiles. The above procedure may then become rather demanding since for every such profile a full propagation with the interface algorithm is required. An alternative way would be then to approximate n_ϕ given ϵ from the start and perform the entire analysis with the backward acceptable window $\omega^\phi_{n,\epsilon}$. If we assume that $D[p(X_{n_\phi}) \| p(X_1)]$ is bounded from above by a known constant M, we find that an approximate value for n_ϕ would satisfy

$$n_\phi \approx n - \left\lfloor \frac{\log(\epsilon / M)}{\log(1 - \delta_A)} \right\rfloor$$

Note that for given ϵ and δ_A, the higher the value of M, the smaller the value of n_ϕ and hence the larger the backward acceptable window. Knowledge of the domain under study can help in determining a suitable value for M. In a medical setting for example, M can be determined by inserting *worst-case scenario* observations for the first time step and computing for that time the posterior probability distribution for the hidden variable from which M can be readily established. The complexity that our method now entails is just the complexity of computing M which is similar to performing a single propagation for a single time step. Note that this computational burden is considerably less than the burden of performing n_ϕ time steps of exact inference, which we thereby forestall in the sensitivity analysis. Note that for some patients the computation of n_ϕ based upon this value

$\Omega = [bc, \bar{b}c, b\bar{c}, \bar{b}\bar{c}]$

$\Omega' = [b, \bar{b}]$



$$S = \begin{bmatrix} b|bc & \bar{b}|bc \\ b|\bar{b}c & \bar{b}|\bar{b}c \\ b|b\bar{c} & \bar{b}|b\bar{c} \\ b|\bar{b}\bar{c} & \bar{b}|\bar{b}\bar{c} \end{bmatrix}$$

Figure 3: The stochastic process B_n depends on the variable C_n. The minimal mixing rate for B_n depends on the stochastic matrix A'. The state spaces before and after the transition are Ω and Ω' respectively.

M will lead to a larger backward acceptable window than the one computed directly from equation (2).

Contraction of multiple subprocesses

In general, a Markovian model with multiple interacting subprocesses can be represented as a single-process stochastic model with a *global* transition matrix A_G by enumerating all combinations of values for the subprocesses. In principle, therefore, we can compute the minimal mixing rate δ_{A_G} for the global matrix and determine an acceptable window as outlined above. Such a procedure, however, is highly time consuming, if not intractable, for models of realistic size. We now show that we can compute a lower bound on δ_{A_G} from knowledge of the contraction rates of the individual subprocesses of the model.

The definition of minimal mixing rate can be generalised to models in which a stochastic subprocess depends not just on its previous state but on the values of some other variables as well. The state space $\Omega = \{w_1, \ldots, w_\nu\}$ before the stochastic transition and the state space $\Omega = \{w'_1, \ldots, w'_{\nu'}\}$ after the transition then are not necessarily the same, and there is an $\nu \times \nu'$ stochastic matrix S rather than a transition matrix A; Figure 3 illustrates the basic idea. Boyen (2002, Theorem 5.11) assumed that a Markovian model could be approximated by conditionally independent sets of subprocesses and that a minimal mixing rate could be computed based on this independence assumption. We now follow a similar approach in establishing a lower bound on δ_{A_G} for any Markovian model.

Theorem *Let $\mathcal{Q}$ be a Markovian model that consists of L subprocesses with stochastic matrices $S_1, \ldots, S_L$, such that each subprocess ℓ depends on at most κ other processes and influences at most q other processes. For each subprocess ℓ, let δ_{S_ℓ} be its minimal mixing rate. Then, a lower bound on the minimal mixing rate δ_{A_G} of the model is*

$$\delta_{A_G} \geq \left(\frac{\min(\delta_{S_1}, \ldots, \delta_{S_L})}{\kappa} \right)^q \cdot \min(\delta_{S_1}, \ldots, \delta_{S_L})^q$$

Proof (sketch): The proof is based on splitting the transition of each subprocess into two consecutive phases, where the first one chooses whether or not to contract, and the second one concludes the transition in a way that depends on whether the subprocess has contracted. Since the two phases form a Markov chain, the mixing rate of $\mathcal{Q}$ is at least that of the first phase alone (Boyen, 2002, Theorem 5.11). A lower bound on the mixing rate for the first phase of a subprocess ℓ that depends on κ other subprocesses now is $\frac{\min(\delta_{S_1}, \ldots, \delta_{S_L})}{\kappa}$.

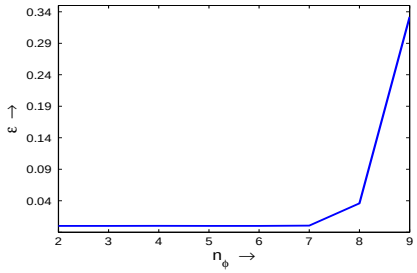


Figure 4: The relationship between n_ϕ and the error ϵ for a specific patient in the dVAP model.

Since the influence of ℓ on another subprocess involves the construction of an intermediate variable for the first phase which contracts independently with rate at least equal to $\min(\delta_{S_1}, \ldots, \delta_{S_L})$, the result in the theorem follows. □

For a Markovian model composed of several sparsely interacting subprocesses each of which is fairly stochastic, we expect a reasonable bound on the overall mixing rate δ_{A_G}. We recall that the larger the mixing rate, the larger the n_ϕ and the smaller the backward window that we can acceptably use for the sensitivity analysis. For the dVAP model, Figure 4 shows, as an example the relationship between the error ϵ and the size of a backward acceptable window for a specific patient. We observe that there is negligible error between the true probability distribution at time step 10 and the one obtained using a value for n_ϕ as high as 7. For all patients in fact, we found that instead of using the observations for all 10 days in the ICU upon performing a sensitivity analysis for the probability of VAP, we can use the observations from day 5 with an average error smaller than $\epsilon = 0.003$. This result is quite very promising for practical reasons since it shows that even if the dynamic processes of a Markovian model are not highly stochastic, the backward acceptable window can still be small enough to allow for good approximations of the sensitivity functions in little time.

Least-square approximation

In general, the aim of performing a sensitivity analysis is to select the parameter probabilities that upon variation show a large effect on the output of the model under study. For this purpose, several concepts have been proposed, such as the concepts of *sensitivity value* (Laskey, 1995) and *admissible deviation* (Van der Gaag and Renooij, 2001); a sensitivity function can be further used to identify changes in the parameter under study that serve to satisfy a query constraint on the output probability (Chan and Darwiche, 2002). These concepts build directly upon the sensitivity functions resulting from the analysis and share that they require further manipulation of these functions.

In (Charitos and Van der Gaag, 2004) we proposed a method to approximate any sensitivity function by a single polynomial of restricted complexity using a *least-squares approximation*. For this purpose, a large number of data points are generated from the established sensitivity func-

tion. Using these data points, estimates are obtained for the coeffi cients of a polynomial f with a desired order that satisfi es the least-squares fi t criterion, where the objective is to minimise

$$Error(f) = \frac{1}{2}\sum_{k}[h_k - f(k)]^2$$

where k and h_k correspond to a point in the interval $[0, 1]$ and its associated value in the true sensitivity function, respectively. The resulting polynomial then is taken as an approximation of the true sensitivity function and used for further manipulation. The order of the approximate function is determined by a *threshold* value for $Error(f)$ which can be established experimentally.

The least-squares approximation technique can be applied not only to exact sensitivity functions, but to the functions obtained using the backward acceptable window as well, thereby providing a *two-stage* approximation of the true sensitivity functions. In this way, we obtain in little time, and without too much loss in accuracy, a single polynomial of relatively low order that describes the influence of the parameter under study on the posterior probability of interest.

As an example, we consider the effect of varying the parameter $\theta = p(rad.signs = yes \mid pneumonia = yes)$ in the dVAP network on the probability of $pneumonia = yes$ at day 10 given the evidence e_{10} for a specifi c patient. The true sensitivity function is a quotient of two polynomials of order 10. Using the backward acceptable window, the resulting approximate sensitivity function is a quotient of polynomials of order 6 each. To simplify this function, we constructed a simpler polynomial as described above. Using 1000 data points generated from the approximate function, we computed a polynomial of order 4. The resulting approximate sensitivity function with respect to θ equals

$$f(\theta) = -6.287{\cdot}\theta^4 + 9.724{\cdot}\theta^3 - 5.08{\cdot}\theta^2 + 0.837{\cdot}\theta + 0.936$$

Figure 5 shows the difference between the exact and the approximate sensitivity functions with or without the least-squares approximation. Note that the two-stage approximation of the true sensitivity function still shows a close fi t to the true sensitivity function.

Conclusions

In this paper, we made a number of contributions to reducing the runtime complexity of sensitivity analysis of Markovian models. We detailed an approximate method for sensitivity analysis that has less runtime requirements than the exact method and yet has a small loss in accuracy. To provide for further computations based upon the approximate sensitivity functions, we presented a method for an additional approximation of their functional form. We illustrated our results using a real-life Markovian model for diagnosing ventilator-associated pneumonia. Our experiments indicate that the sensitivity functions for our model can be computed effi ciently with just minor fluctuations from their exact values. In the future, we plan to perform additional experiments to support our current results and also to study the joint influence of two parameters on the output probability of interest.

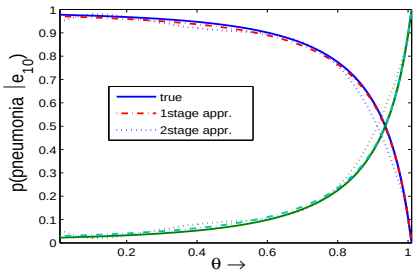


Figure 5: Comparison of the exact and approximate sensitivity functions expressing the probabilities $p(pneumonia = yes)$ (upper set of plots) and $p(pneumonia = no)$ (bottom set of plots) given e_{10} in terms of the parameter $\theta = p(rad.signs = yes \mid pneumonia = yes)$.

Acknowledgements

This research was (partly) supported by the Netherlands Organization for Scientifi c Research (NWO).

References

Boyen, X. 2002. *Inference and Learning in Complex Stochastic Processes*. Ph.D. diss, Stanford University.

Chan, H. and Darwiche, A. 2002. When do numbers really matter? *Journal of Artificial Intelligence Research*, 17:265-287.

Charitos, T. and Van der Gaag, L.C. 2004. Sensitivity properties of Markovian models. *Proceedings AISTA Conference in Cooperation with IEEE Computer Society*, ISBN: 2-9599776-8-8.

Charitos, T., Van der Gaag, L.C., Visscher, S., Schurink, K. and Lucas, P. 2005. A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. Working notes of the 10th Workshop on Intelligent Data Analysis in Medicine and Pharmacology, pp. 32-37.

Coupé, V.M.H. and Van der Gaag, L.C. 2002. Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence* 36:323-356.

Kjaerulff, U. and Van der Gaag, L.C. 2000. Making sensitivity analysis computationally effi cient. *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 317-325.

Laskey, K.B. 1995. Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics* 25:901-909.

Murphy, K.P. 2002. *Dynamic Bayesian Networks: Representation, Inference and Learning*. Ph.D. diss, University of California Berkley.

Rabiner, L.R. 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257-286.

Van der Gaag, L.C. and Renooij, S. 2001. Analysing sensitivity data from probabilistic networks. *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 530-537.