



## Bayesian network modeling of correlated random variables drawn from a Gaussian random field

Michelle Bensi<sup>a,\*</sup>, Armen Der Kiureghian<sup>a</sup>, Daniel Straub<sup>b</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, United States

<sup>b</sup> Engineering Risk Analysis Group, Technische Universität München Arcistr. 21, 80290 München, Germany

### ARTICLE INFO

#### Article history:

Received 26 August 2010

Received in revised form 1 May 2011

Accepted 3 May 2011

Available online 19 October 2011

#### Keywords:

Bayesian network

Correlation matrix

Infrastructure systems

Probabilistic updating

Random fields

Spatially distributed systems

### ABSTRACT

In many civil engineering applications, it is necessary to model vectors of random variables drawn from a random field. Furthermore, it is often of interest to update the random field model in light of available or assumed observations on the random field or related variables. The Bayesian network (BN) methodology is a powerful tool for such updating purposes. However, there is a limiting characteristic of the BN that poses a challenge when modeling random variables drawn from a random field: due to the full correlation structure of the random variables, the BN becomes densely connected and inference can quickly become computationally intractable with increasing number of random variables. In this paper, we develop approximation methods to achieve computationally tractable BN models of correlated random variables drawn from a Gaussian random field. Using several generic and systematic spatial configuration models, numerical investigations are performed to compare the relative effectiveness of the proposed approximation methods. Finally, the effects of the random field approximation on estimated reliabilities of example spatially distributed systems are investigated. The paper concludes with a set of recommendations for BN modeling of random variables drawn from a random field.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

In civil engineering applications, it is often necessary to model vectors of random variables drawn from a random field. For example, in investigating the seismic risk of a lifeline, the earthquake-induced ground motion intensities at the locations of the system components constitute a vector of random variables drawn from the ground motion random field. Similarly, factors determining the progress of deterioration in elements of concrete surfaces are random variables drawn from environmental and material property random fields. Proper modeling of the dependence structure of vectors of random variables is essential for accurate probabilistic analysis. In the special case when the field is Gaussian, or derived from a Gaussian field, the spatial dependence structure of the field is completely defined by the autocorrelation function and the correlation matrix fully defines the dependence structure of the random vector drawn from the field. Typically, this correlation matrix is fully populated. Although this paper only deals with Gaussian random fields, the methods developed are equally applicable to non-Gaussian fields that are derived from Gaussian fields, e.g., [1].

In some applications, including the aforementioned examples, it is of interest to update a probabilistic model in light of available or assumed observations of the random field. For example, in the case of a lifeline subjected to an earthquake, one might be interested in updating the reliability of the system when ground motion intensities at one or more locations are observed, or when evidence is available on the performance of individual components based on the output from structural health monitoring sensors or observations made by inspectors [2]. In the case of a concrete surface subject to deterioration, the reliability of the system can be updated, e.g., when cracking (or no cracking) of the concrete in some of the elements is observed. The Bayesian network (BN) methodology is a powerful tool for such updating purposes, particularly when the available information evolves in time and the updating must be done in (near) real time, see, e.g., [3,4]. However, there is a limiting characteristic of the BN that poses a challenge when modeling random variables drawn from a random field: due to the full correlation structure of the random variables, the BN becomes densely connected. When combining these random variables with system models that involve additional random variables, the computational and memory demands of the resulting BN rapidly grow with the number of points drawn from the random field. In this paper, we develop approximate methods to overcome this difficulty. Specifically, we present methods for reducing the density of the BN model of the random field by selectively eliminating nodes and

\* Corresponding author.

E-mail addresses: [mtbensi@gmail.com](mailto:mtbensi@gmail.com) (M. Bensi), [adk@ce.berkeley.edu](mailto:adk@ce.berkeley.edu) (A. Der Kiureghian), [straub@tum.de](mailto:straub@tum.de) (D. Straub).

links. The aim is to minimize the number of links in the BN while limiting the error in the representation of the correlation structure of the random variables drawn from the Gaussian random field.

When the random field as well as the observed random variables are jointly Gaussian, a well-known analytical solution exists for computing the conditional probabilistic model. However, the random field model often is part of a larger problem involving mixtures of continuous and discrete random variables and fields. For example, in seismic risk assessment of a lifeline, a random field may define the ground motion intensity across a geographic region, while discrete random variables define the performance or damage states of the lifeline and its constituent components. When evidence is entered on non-Gaussian or discrete random variables in such a model, e.g., the observed damage state of a component, the existing analytical solution for updating the distribution of the Gaussian variables is no longer applicable. It is in this context that the BN is useful for modeling and updating of Gaussian random fields.

The paper begins with a brief introduction to BNs as a means for probabilistic inference and describes their advantages and limitations. Next, BN models of random variables drawn from a Gaussian random field are described. Approximation methods are then developed to achieve computationally tractable BN models. Using several generic and systematic spatial configuration models, numerical investigations are performed to compare the relative effectiveness of the proposed approximation methods. Finally, the effects of the random field approximation on estimated reliabilities of example spatially distributed systems are investigated. The paper ends with a set of recommendations for BN modeling of random variables drawn from a random field. More details on development of BN models for random fields and application to infrastructure seismic risk assessment can be found in [2].

## 2. Brief Introduction to Bayesian network

Bayesian networks are probabilistic graphical models consisting of nodes representing random variables and directed links describing probabilistic dependencies. Throughout the paper, the terms “node” and “random variable” are used interchangeably. Consider the simple two-node BN in Fig. 1a, which models random variables  $X$  and  $Y$ . The directed link indicates that  $Y$  is probabilistically dependent on  $X$ . Node  $Y$  is a *child* of node  $X$ , while node  $X$  is a *parent* of node  $Y$ . Attached to node  $Y$  is the conditional probability distribution of  $Y$  given  $X$ . Because node  $X$  has no parent, a marginal probability distribution is attached to it. If the two variables are discrete, then probability mass functions (PMFs) define their distributions. In particular, a conditional PMF defines the probability that  $Y$  is in each of its mutually exclusive states, given each mutually exclusive state of  $X$ , i.e., the probabilities  $\Pr(Y = y_i | X = x_j)$ , where  $y_i$  indicates the  $i$ th state of  $Y$  and  $x_j$  indicates the  $j$ th state of  $X$ . In the BN terminology, this conditional PMF is called a *conditional probability table* (CPT). If node  $Y$  has multiple parents, as in Fig. 1b, the size of the CPT for node  $Y$  becomes large because the

PMF of  $Y$  must be defined for all combinations of the states of the parent nodes  $X_1, \dots, X_n$ . If each of the nodes in Fig. 1b has  $m$  states, then the CPT attached to node  $Y$  has  $m^{n+1}$  entries. It is seen that the size of the CPT attached to a node grows exponentially as the number of parents increases. These CPTs generally must be stored in memory. Therefore, as  $n$  increases, computational bottlenecks are encountered due to physical memory constraints.

BNs are most useful in answering probabilistic queries, e.g., determining the posterior distributions of the random variables in the BN, when one or more variables are observed. The process of updating the BN given available evidence is known as probabilistic inference. Updating occurs consistent with the  $d$ -separation properties of the BN, which characterize the way in which information flows through different types of connections (see [5,6] for more details). Conceptually, inference may be thought of as efficient application of Bayes Rule and the Total Probability Theorem on a large scale.

Many applications in civil engineering require mixtures of continuous and discrete nodes. For example, as described earlier, in seismic applications the distribution of the ground motion intensity at a site is modeled by a continuous random variable, while the damage state of a component at that location may be represented by a discrete random variable. However, existing exact inference algorithms for BNs and software applications utilizing these algorithms [7,8] impose severe limitations on the use of continuous random variables. Specifically, they only permit linear functions of Gaussian random variables that have no discrete children, for which case analytical solutions exist for computing the conditional probability distribution [9]. The necessity of modeling discrete children of continuous nodes, e.g., the damage state of a component that is causally dependent on the ground motion intensity at its site, prohibits the use of the aforementioned algorithms. Furthermore, inference algorithms for BNs with mixtures of discrete and continuous random variables are computationally far more demanding than those for BNs with only discrete nodes [10,11]. Hence, when using BNs, there is often significant advantage in discretizing all continuous random variables. The examples utilized in this paper are solved using exact inference algorithms that require all nodes to be discretized.

One of the most common exact inference algorithms is the Junction Tree algorithm [5,12]. When performing exact probabilistic inference using this algorithm, graphical constructs known as *cliques* are formed, which contain subsets of nodes in the BN. Each clique is assigned a *clique table*, created by taking the product of the CPTs of all nodes in the clique. Common general-purpose BN software, such as Hugin [7,13], require that clique tables formed during probabilistic inference be stored in memory. For BNs with discrete nodes, the memory demand of the Junction Tree algorithm is exponential in the size of the largest clique generated when performing inference. Therefore, it is preferable to work with cliques, and consequently clique tables, that are as small as possible. The sizes of cliques generated when performing inference calculations are related to the sizes of the CPTs associated with nodes in the graph as well as the density of dependency (links) between the nodes. Thus, discrete-node BNs with broadly dependent random variables (i.e., densely connected nodes) and/or nodes with many states (i.e., large CPTs) result in large cliques and, consequently, exponential increases in computational demands. The reader is referred to common texts, e.g., [5], for more details on clique sizes and the Junction Tree algorithm. Although our interest is in the application of exact inference algorithms, it is noted that small CPTs are also preferable when working with approximate sampling-based algorithms, e.g., [14,15]. Therefore, by reducing CPT sizes, the methods described in this paper are also useful when applying approximate inference algorithms.

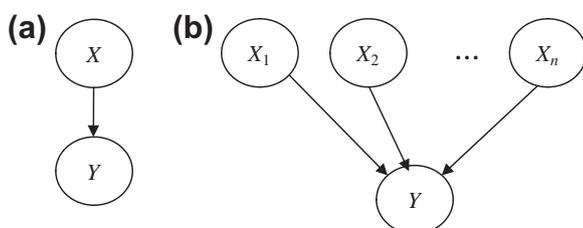


Fig. 1. (a) Two-node BN with  $Y$  dependent on  $X$  and (b) BN with  $Y$  dependent on a vector of random variables  $X = \{X_1, \dots, X_n\}$ .

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات