www.elsevier.com/locate/ress

Modelling dependable systems using hybrid Bayesian networks

Martin Neil, Manesh Tailor, David Marquez, Norman Fenton, Peter Hearty

PII: S0951-8320(07)00104-4
DOI: doi:10.1016/j.ress.2007.03.009
Reference: RESS 3870

To appear in: *Reliability Engineering and System Safety*

Cite this article as: Martin Neil, Manesh Tailor, David Marquez, Norman Fenton and Peter Hearty, Modelling dependable systems using hybrid Bayesian networks, *Reliability Engineering and System Safety* (2007), doi:10.1016/j.ress.2007.03.009

Modelling Dependable Systems using Hybrid Bayesian Networks

Martin Neil[†‡], Manesh Tailor[‡], David Marquez[†], Norman Fenton[†‡] and Peter Hearty[†]
† Department of Computer Science, Queen Mary, University of London
‡ Agena Ltd

Abstract

A hybrid Bayesian Network (BN) is one that incorporates both discrete and continuous nodes. In our extensive applications of BNs for system dependability assessment, the models are invariably hybrid and the need for efficient and accurate computation is paramount. We apply a new iterative algorithm that efficiently combines dynamic discretisation with robust propagation algorithms on junction tree structures to perform inference in hybrid BNs. We illustrate its use in the field of dependability with two example of reliability estimation. Firstly we estimate the reliability of a simple single system and next we implement a hierarchical Bayesian model. In the hierarchical model we compute the reliability of two unknown subsystems from data collected on historically similar subsystems and then input the result into a reliability block model to compute system level reliability. We conclude that dynamic discretisation can be used as an alternative to analytical or Monte Carlo methods with high precision and can be applied to a wide range of dependability problems.

Keywords: Bayesian Networks; Bayesian Software; Systems Dependability; Dynamic Discretisation.

1. Introduction

We have used Bayesian Nets (BNs) in a range of real-world applications of system dependability assessment (see for example [15, 16, 3]). In such applications it is inevitable that there will be a mixture of discrete and continuous nodes (the resulting BNs are called hybrid). The traditional approach to handling (non-Gaussian) continuous nodes is static: you have to discretise them using some pre-defined range and

intervals. However, this approach is unacceptable for critical type systems where there is a demand for reasonable accuracy. To overcome this problem we have developed a new and powerful approximate algorithm for performing inference in hybrid BNs. We use a process of *dynamic discretisation* of the domain of all continuous variables in the BN. The approach is influenced by the work of Kozlov and Koller [9] using entropy error as the basis for approximation. We differ from their approach by integrating an iterative approximation scheme within existing BN software architectures, such as in Junction Tree (JT) propagation [7]. Thus, rather than support separate data structures and a new propagation algorithm we use the data structures commonly used in JT algorithms.

The power and flexibility of the approach is demonstrated by applying it to estimate the reliability of repairable systems represented by a Bayesian hierarchical model. This problem represents a very simplified version of fragments of the wide range of models we have implemented as part of commercial and research projects. These have been in areas as diverse as data fusion, parameter learning, discrete systems simulation, RAM (Reliability, Availability and Maintainability) evaluation and software defect prediction. The modelling has been made possible because our dynamic discretisation algorithm has recently been implemented in the commercial general-purpose Bayesian Network software tool AgenaRisk [1].

2. Background

BNs have been widely used to represent full probability models in a compact and intuitive way. In the BN framework the independence structure in a joint distribution is characterised by a directed acyclic graph, with nodes representing random variables (which can be discrete or continuous, and may or may not be observable), and directed arcs representing causal or influential relationship between variables [18]. The conditional independence assertions about the variables, represented by the lack of arcs, reduce significantly the complexity of inference and allow the underlying joint probability distribution to be decomposed as a product of *local conditional probability distributions* (*CPD*) associated with each node and its respective parents. If the variables are discrete, the CPDs can be represented as Node Probability Tables (NPTs), which list the probability that the child node takes on each of its different values for each combination of values of its parents. Since a BN encodes all relevant qualitative and quantitative

information contained in a full probability model, it is an excellent tool for many types of probabilistic inference where we need to compute the posterior probability distribution of some variables of interest (unknown parameters and unobserved data) conditioned on some other variables that have been observed.

A range of robust and efficient propagation algorithms has been developed for exact inference on Bayesian networks with discrete variables [11, 20, 7]. The common feature of these algorithms is that the exact computation of posterior marginals is performed through a series of local computations over a secondary structure, a tree of clusters, enabling calculation of the marginal without computing the joint distribution. See also [5].

The present generation of BN software tools attempt to model continuous nodes by numerical approximations using static discretisation as implemented in a number of software tools [6, 14]. Although discretisation allows approximate inference in a hybrid BN without limitations on relationships among continuous and discrete variables, current software implementations require users to define a uniform discretisation of the states of any numeric node (whether it is continuous or discrete) as a sequence of pre-defined intervals, which remain *static* throughout all subsequent stages of Bayesian inference regardless of any new conditioning evidence. The more intervals you define, the more accuracy you can achieve, but at a heavy cost of computational complexity. This is made worse by the fact that you do not necessarily know in advance where the posterior marginal distribution will lie on the continuum for all nodes and which ranges require the finer intervals. It follows that where a model contains numerical nodes having a potentially large range, results are necessarily only crude approximations.

Alternatives to discretisation have been suggested by [12, 2], who describe potential approximations using mixtures of truncated exponential (MTE) distributions, [8] who combine MTE approximations with direct sampling (Monte Carlo) methods, and [13] who uses variational methods. There have also been some attempts for approximate inference on hybrid BNs using Markov Chain Monte Carlo (MCMC) approaches [19]. However, constructing dependent samples that mixed well (i.e., that move rapidly throughout the support of the target posterior distribution) remains a complex task.

3. Dynamic Discretisation

Let X be a continuous random node in the BN. The range of X is denoted by Ω_X, and the probability density function (PDF) of X, with support Ω_X, is denoted by f_X. The idea of discretisation is to approximate f_X as follows:

1. Partition Ω_X into a set of interval $\Psi_X = \{w_j\}$, and
2. Define a locally constant function $\tilde{f}_X$ on the partitioning intervals.

Discretisation operates in much the same way when X takes integer values but in this paper we will focus on the case where X is continuous.

As Kozlov and Koller [9], we use an upper bound of the Kullback-Leibler (KL) metric between two density functions f and g:

$$D(f \| g) = \int_S f(x) \log \frac{f(x)}{g(x)} dx$$

as an estimate of the relative entropy error induced by the discretised function. Under the KL metric, the optimal value for the discretised function $\tilde{f}$ is given by the mean of the function f in each of the intervals of the discretised domain. The main task reduces then to finding an optimal discretisation set $\Psi_X = \{\omega_j\}$.

Our approach to dynamic discretisation searches Ω_X for the most accurate specification of the high-density regions given the model and the evidence, calculating a sequence of discretisation intervals in Ω_X iteratively. At each stage in the iterative process, a candidate discretisation, Ψ_X, is tested to determine

whether the relative entropy error of the resulting discretised probability density $\tilde{f}_X$ is below a given threshold, defined according to the trade off between the acceptable degree of precision and computation time.

By dynamically discretising the model we achieve more accuracy in the regions that matter and incur less storage space over static discretisations. Moreover, we can adjust the discretisation any time in response to new evidence to achieve greater accuracy. In outline, dynamic discretisation follows these steps:

1. Convert the BN to a junction tree (JT) and choose an initial discretisation for all continuous variables.
2. Calculate the NPT of each node given the current discretisation
3. Enter evidence and perform global propagation on the junction tree, using standard JT algorithms [7].
4. Query the BN to get posterior marginals for each node, compute the approximate relative entropy error, and check if it satisfies the convergence criteria.
5. If not, create a new discretisation for the node by splitting those intervals with highest entropy error.
6. Repeat the process by recalculating the NPTs and propagating the BN, and then querying to get the marginals and then split intervals with highest entropy error.
7. Continue to iterate until the model converges to an acceptable level of accuracy.

4. Estimating System Reliability using Bayes

In this section we illustrate the efficacy of the dynamic discretisation approach to estimate the reliability of a continuous use repairable system. Bayesian modelling offers a suitable framework for assessing the reliability of such systems, allowing us to integrate information from multiple sources at different levels of

granularity, as well as expert opinion. Here we are interested in assessing if a repairable system is meeting its reliability requirement for Mean Time Between Failure (MTBF). The model considered here is based on the following assumptions:

1. Repairs take place instantaneously after failure, and the repair times are assumed to be negligible. Furthermore, the waiting time between successive failures are independent, identically distributed exponential random variables with constant failure rate, λ, assumed unknown. For repairable systems with negligible repair times, this implies that the Mean Time Between Failure, *MTBF*, is near the Mean Time To Failure, which, for the exponential model, is given by $1/\lambda$. The cumulative number of failures up to time *t*, $r = N(t)$, can therefore be modelled as a Homogeneous Poisson Process (HPP), with mean given by λt, the expected number of failures up to time *t*. That is,

$$p\left(N(t) = r \middle| \lambda\right) = Poisson(\lambda t) \propto (\lambda t)^r e^{-\lambda t}$$

2. The uncertainty about the failure rate, λ, is modelled by a prior probability distribution, $p(\lambda)$, representing our previous information or judgment about the system's reliability. Because of their mathematical convenience, conjugate priors are used extensively in Bayesian modelling [4]. For the Poisson distributed number of failures, $N(t)$, the most appropriate choice for the distribution of λ is the gamma distribution:

$$p\left(\lambda \middle| a, b\right) = Gamma(a, b) \propto \lambda^{a-1} e^{-b\lambda}$$

with the parameters *a* and *b* chosen according to our prior knowledge [17].

Under the above assumptions, after a test is run on a given subsystem during *T* operating hours, the observed number of failures, $N(T) = r$, is combined with the prior distribution of λ to produce the updated or posterior distribution model for λ:

$$
\begin{aligned}
p(\lambda|r) &\propto p(N(T)=r \mid \lambda)\, p(\lambda|a,b) \\
&= Poisson(\lambda T)\, Gamma(a,b) \\
&= Gamma(a+r, b+T)
\end{aligned}
$$

With dynamic discretisation we do not need to assume conjugacy; hence the prior distribution for the rate could be any sensible shape and scale (a non-conjugate example, that uses an expert opinion as a prior, is provided with the evaluation version of AgenaRisk). However, to aid comparison, we will first show a simple example of a single-component system using a conjugate prior. The BN model is depicted in Figure 1. A noninformative Gamma prior distribution for the unknown failure rate, λ, was assumed by setting $p(\lambda|a,b) = Gamma(0.001, 0.001)$. In addition to nodes for each of the variables defined above, we have added an additional discrete node to test the requirement that the system will survive for more than 600 hours, thus transforming it to a hybrid BN. In Figure 1 the node labels are annotated with their relevant functions.

If we observe five failures in 2000 hours of use, the estimated marginal distributions after 50 iterations in AgenaRisk are shown in Figure 2. The mean failure rate estimate, $E(\lambda)$, is 0.00249 and the probability of surviving 600 hours is 0.244. This compares very well with the analytical solution, $E(\lambda|r,T) = 0.0025$.

5. Estimating System Reliability using a Hierarchical Bayesian model

Let us now consider a system comprised of two subsystems, *k* = 1, 2, but we have not as yet been able to directly test the reliability of each one. Instead we can estimate the reliability of each subsystem from historical data gathered on three similar subsystems, C_{jk}, with $j = 1,2,3$, drawn from the same family as the subsystem of interest. Thus the unseen subsystem is considered to be exchangeable with the historical subsystems. For example the subsystem of interest might be a 2 litre engine and the historical data might be on a 3 litre, 1.6 litre and a 1 litre engine.

There are thus six failure rates to estimate, three for each of the historical subsystems and from these we wish to estimate the posterior predictive reliability distribution for each subsystem of interest.

In order to assess the reliability of the historical subsystems, we assume that a series of reliability tests have been conducted under the same operational settings. The data resulting from these tests consist in the number of failures observed after a fixed period of testing time. Thus, if *n* independent tests were conducted on each historical subsystem for a defined period of time T, these results in *n* independent failure counts, with underlying Poisson population distribution:

$$\left\{N_{ij}\right\}_{i=1}^{n} \overset{iid}{\sim} \text{Poisson}\left(\lambda_j T\right), \quad j=1,2,3.$$

The unknown failure rates, $\left\{\lambda_j\right\}$, of the similar subsystem are assumed exchangeable in their joint distribution, reflecting the lack of information – other than data – about the reliability of the subsystems. The parameters $\left\{\lambda_j\right\}$ are thus considered a sample from the conjugate gamma prior distribution, governed by unknown hyperparameters *a*, *b*:

$$p\left(\lambda_j / a,b\right) = Gamma\left(a,b\right)$$

To complete the hierarchical specification, we need to assign a prior probability distribution to the hyperparameters *a*, *b*. Since no joint conjugate prior is available when *a*, *b* are both assumed unknown, their prior distributions are specified independently. If no information is available about the hyperparameters, we can assign them vague prior distributions, for example, by defining $p\left(a,b\right) = p\left(a\right)p\left(b\right) = \exp(0.1)\exp(0.1)$. However, because reliability data can be sparse and heavily censored, additional information in the form of expert judgement plays an important role in the definition of statistical reliability models. Here we choose illustrative distributions for shape and scale parameters, but in practice these might be elicited by experts. For example here we use:

$$a \sim Triangular\left(0,10,100\right) \text{ and } \log_{10} b \sim Triangular\left(-6,-3,-1\right)$$

There are two points to make about the example distributions we have used for the hyperparameters. They are not conjugate deliberately to illustrate the flexibility of our approach. Secondly, the prior is based on past experience, and in this particular case:

a) asking experts to use a triangular distribution is relatively easier compared to using other more complex distributions, and

b) the parameters, (a,b), can be interpreted in terms of failure counts and exposures respectively. So $Triangular(0,10,100)$ has a modal failure count around 10 and decreasing probability of experiencing up to 100 failures, and this might be the range of values observed in past practice. Similarly the exposures are typically in the region 10^3 with diminishing probability as we approach 10^6.

Test data for assessing the reliability of the subsystems *k* = 1,2 are provided in

Table1. Eight independent tests were conducted, under the same operational environment, for each of the corresponding interchangeable subsystem C_{jk}, with $j = 1,2,3$. The numbers of failures observed after 2000 hours of exposure are depicted in the table.

The BN model for the reliability of the subsystem 1 is depicted in Figure 3, with the marginal probability distributions superimposed on the respective nodes.

The labelled nodes at the bottom of the graph represent the observed number of failures after 2000 hours of exposure test for each of the historical subsystems, as depicted in

Table 1. As illustrated in the graph, sample Poisson distributions depend on the failure rates for the respective subsystem (C11, C21, C31), represented by the intermediate nodes, which in turn are assumed to be sampled from a gamma population distribution with hyperparameters *a* = shape and *b* = scale.

In general, multi-parameter models do not allow simple explicit calculation of posterior distributions. The dynamic discretisation algorithm presented in this paper allows us to compute the posterior distributions of

the system reliability in a straightforward way. The Figure 3 shows the conditional posterior distribution of the failure rates, after observing the failures counts resulting from the tests. Namely,

$$P\left(\lambda_j/(a,b),\{N_{ij}\}_i\right) = Gamma\left(a+\sum_{i=1}^{n} N_{ij}, b+nT\right)$$

Similarly, Figure 3 shows the non-standard marginal posterior distributions for the hyperparameters, $P\left(a/\{N_{ij}\}\right), P\left(b/\{N_{ij}\}\right)$, which express our uncertainty about the parameters *a* and *b*, after taking into account the information provided by the observed Poisson counts. From these posterior distributions, we can obtain an estimate of the overall reliability of the subsystem of interest, as expressed by the conditional posterior distribution $P\left(\lambda/(a,b),\{N_{ij}\}_i\right)$, which allows us to estimate the overall reliability for the MTBF of each subsystem. For the given data, the predicted mean failure rate for subsystem 1 is $E\left[\lambda|a,b,\{N_{ij}\}\right]$ = 0.0018 and the predicted mean for the MTBF is 616.25. Analogously, we build a BN for the subsystem 2, and obtain estimates for its overall reliability, $E\left[\lambda|a,b,\{N_{ij}\}\right]$ = 0.00081. The predicted mean for the MTBF is 2176.

The resulting the posterior distribution of the overall reliabilities for the MTBF of each subsystem become the ancestor nodes of a reliability block diagram for the entire system. This is depicted in Figure 4.

The discrete intermediate nodes are used to test the requirement that each subsystem will survive for a given number of hours, given the estimated overall reliability of each subsystem. For the present example, we obtain that the probability of subsystem 1 surviving 600 hours is 0.3592, and the probability of subsystem 2 surviving 600 hours is 0.9852. The top node in the BN determines the probability of the joint event that both subsystems satisfy their corresponding reliability requirement for the MTBFs. For

independent subsystems, this is given by the product of the probabilities of the respective events, resulting in a probability of the system surviving of 0.35395

6. Concluding Remarks

We have provided an overview of a new approximate inference algorithm designed for a general class of hybrid BNs. This dynamic discretisation algorithm (implemented in the AgenaRisk software) finally frees BN modellers from the burden (and inaccuracies) associated with having to statically discretise continuous nodes. We have first highlighted how this approach enables us to estimate reliability of a simple system, for which the results compare very favourably with analytical methods. We then assess the reliability of a more complex system comprised of two subsystems using a Bayesian hierarchical model, which allows the integration of data and expert opinion, available at different levels. The most common estimation strategy for such hierarchical models, where the resulting joint distribution of the associated model parameters cannot be evaluated analytically, has been to use intensive sampling algorithms, collectively known as Markov Chain Monte Carlo (MCMC) methods, from which approximate solutions can be obtained after drawing probably ten of thousand of dependent samples. We have shown how our combined scheme of dynamic discretisation and robust propagation algorithms on hybrid BNs can be used to obtain accurate results, offering a viable alternative to MCMC approaches, implemented within an easy-to-use and user friendly environment.

It is a simple leap from this example to considerably more complex examples, say involving families of systems modelled hierarchically or using censored data, where

dynamic discretisation could provide alternative and perhaps better solutions to those provided by other approximate methods such as MCMC. Typical solutions to this involve a variety of complex algorithms not found in discrete BN packages, yet a solution is easily and quickly produced using dynamic discretisation.

Acknowledgments

We would like to extend our thanks to the referees for their insightful comments and helpful suggestions. This paper is based in part on work undertaken on the eXdecide: "Quantified Risk Assessment and Decision Support for Agile Software Projects" project, EPSRC project code EP/C005406/1, March 2005 - Feb 2008.

References

[1]Agena Ltd. 2005, www.agenarisk.com

[2] Cobb B, Shenoy P. Inference in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials, University of Kansas, School of Business, working paper 294. 2003.

[3] Fenton NE, Marsh W, Neil M, Cates P, Forey S, Tailor T, Making Resource Decisions for Software Projects", 26th Intl Conf on Software Engineering (ICSE 2004), IEEE Comp Society 2004, ISBN 0-7695-2163-0, p. 397-406.

[4] Gelman A, Carlin J. B, Stern H. S, Rubin D. B. Bayesian Data Analysis (2nd Edition), Chapman and Hall, 2004. p. 209-302.

[5] Huang C, Darwiche A. Inference in belief networks: A procedural guide. Int. J. Approx. Reasoning 15(3). 1996. p. 225-263

[6] Hugin. www.hugin.com. 2005.

[7] Jensen F, Lauritzen S.L, Olesen K. Bayesian updating in recursive graphical models by local computations, Computational Statistics Quarterly, 4. 1990. p. 260-282

[8] Koller D, Lerner U, Angelov D. A general algorithm for approximate inference and its applications to Hybrid Bayes Nets, in K.B Laskey and H. Prade (eds.), Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence. 1999. p. 324-333.

[9] Kozlov A.V, Koller D. Nonuniform dynamic discretization in hybrid networks, in D. Geiger and P.P. Shenoy (eds.), Uncertainty in Artificial Intelligence, 13. 1997. p. 314-325.

[10] Lauritzen S.L., and Jensen F. Stable local computation with conditional Gaussian Distributions, Statistics and Computing, 11. 2001. p. 191-203.

[11] Lauritzen S.L, Speigelhalter D.J. Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion), Journal of the Royal Statistical Society Series B, Vol. 50, No 2. 1988. p.157-224.

[12] Moral S, Rumı R, Salmeron A. Mixtures of truncated exponentials in hybrid Bayesian networks, in P. Besnard and S. Benferhart (eds.), Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Lecture Notes in AI, 2143. 2001. p. 156-167.

[13] Murphy K. A variational approximation for Bayesian networks with discrete and continuous latent variables, in Laskey and H. Prade (eds.), Uncertainty in AI, 15. 1999. p. 467-475.

[14] Netica. www.norsys.com. 2005.

[15] Neil M, Fenton N, Forey S, Harris R. Using Bayesian Belief Networks to Predict the Reliability of Military Vehicles, IEE Computing and Control Engineering J 12(1). 2001. p. 11-20

[16] Neil M, Malcolm B, Shaw R. Modelling an Air Traffic Control Environment Using Bayesian Belief Networks, 21st International System Safety Conference, Ottawa, Ontario, Canada. 2003a.

[17] NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/. 2006.

[18] Pearl J. Graphical models, causality, and intervention, Statistical Science, vol 8, no. 3. 1993. p. 266-273

[19] Shacter R, Peot M. Simulation approaches to general probabilistic inference on belief networks. In Proceedings of the 5th Annual Conference on Uncertainty in AI (UAI). 1989. p. 221-230.

[20] Shenoy P, Shafer G. Axioms for probability and belief-function propagation, Readings in uncertain reasoning, Morgan Kaufmann Publishers Inc. 1990. p. 575-610 .

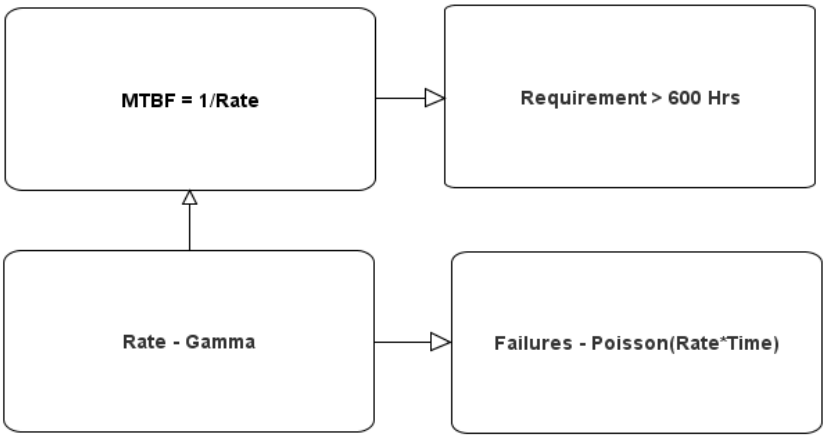

Figure 1 BN for reliability estimation example

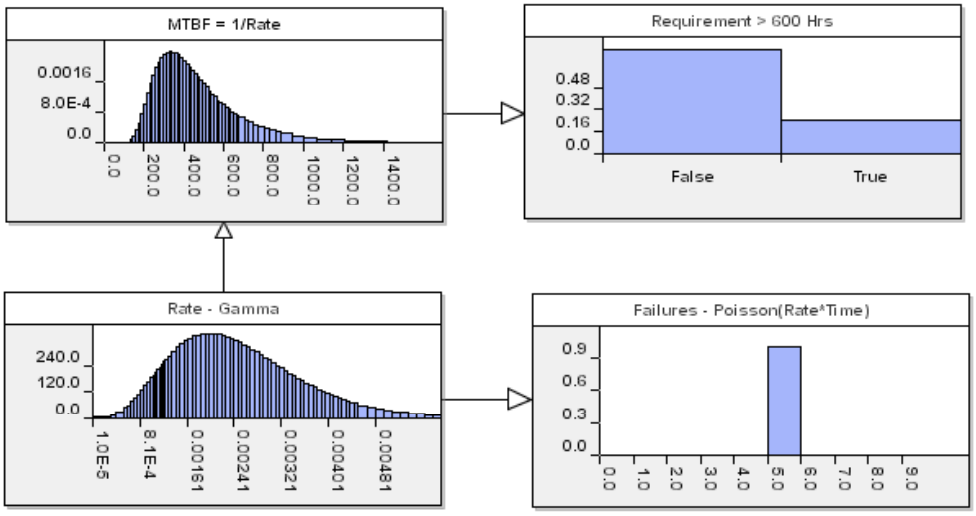

Figure 2 BN for reliability estimation example with marginal distributions superimposed on the graph

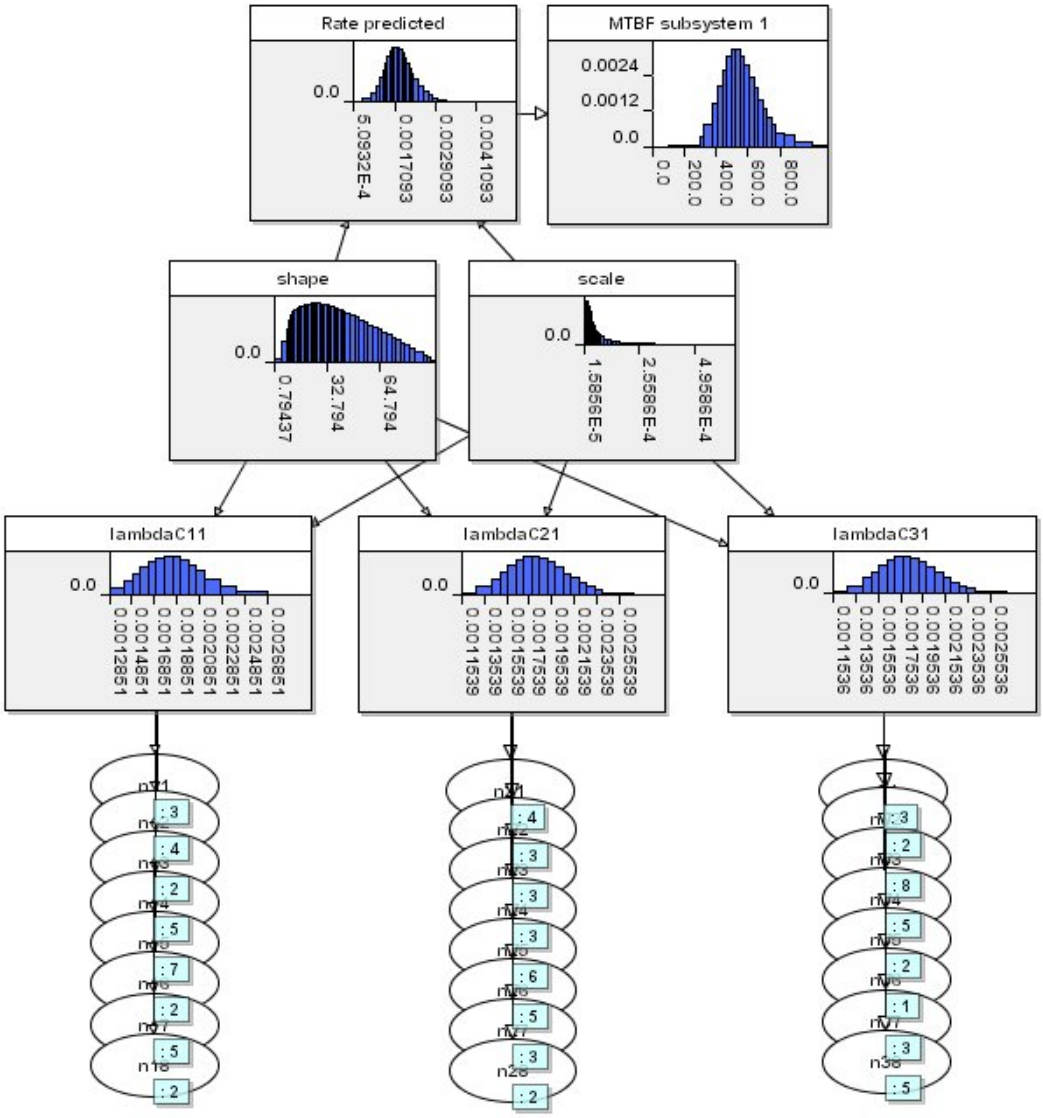

Figure 3 BN for reliability estimation of subsystem 1. Marginal distributions superimposed on the graph

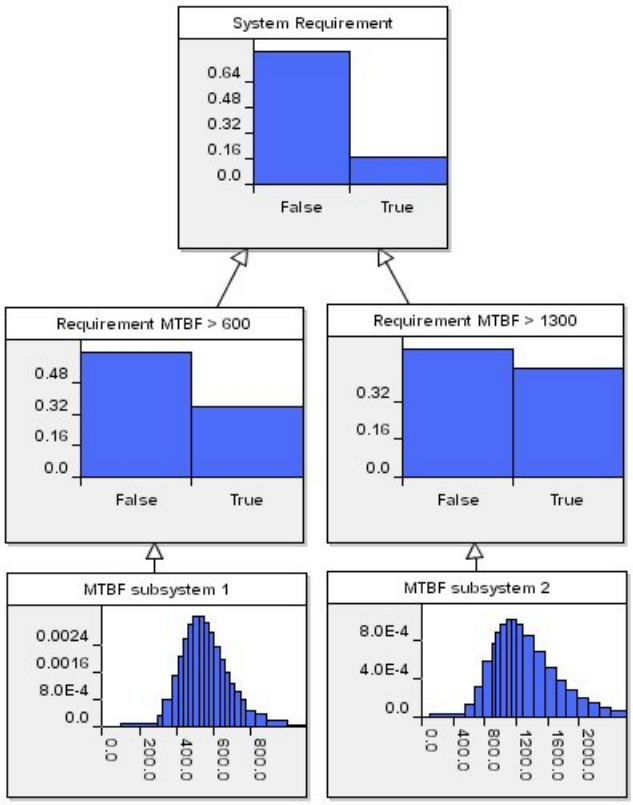

Figure 4 BN modelling a reliability block diagram for the system. Marginal distributions superimposed on the graph.

Table 1: Test data

Subsystem	Data
C11	3 4 2 5 7 2 5 2
C21	4 3 3 3 6 5 3 2
C31	3 2 8 5 2 1 3 5
C12	0 1 1 0 3 1 0 1
C22	1 2 2 1 3 2 1 2
C32	4 3 3 3 1 1 0 2