Modelling and Inference for Networks with Repairable Redundant Subsystems

Louis J.M. Aslett and Simon P. Wilson

School of Computer Science and Statistics
Trinity College Dublin, Ireland
E-mail: louis@maths.tcd.ie / swilson@tcd.ie

Abstract

We consider the problem of modelling the reliability of a network of subsystems where each subsystem has redundancy and is repairable. The motivation for this work is large-scale telecommunications networks.

The time to failure of the subsystem hardware is modeled by an appropriate Markov process and is hence a phase-type distribution. The network structure defines a failure rule in terms of the states of the subsystems, allowing computation by Monte Carlo simulation of the time to failure distribution for the network. This is illustrated by some simple examples. When data on the reliability of the subsystems are available, this can be incorporated via a Bayesian inference approach to modify the prediction of network reliability.

1 Introduction

It is extremely common to find redundant subsystems in mission critical networks, where it is possible to perform non-disruptive repairs in the event of covered failures. The fact that repairs can take place without failure means that traditional failure rate distributions such as the Weibull and exponential can be a poor fit in such situations. Moreover, they fail to reflect the inherent stochastic process underlying the repairable subsystem.

This situation has been successfully modelled using phase-type distributions (for example Neuts et al. (2000)). However, there has been limited work in the literature seeking to model subsystem redundancy in the context of the system and network in which it resides, as well as little emphasis on statistical inference. In particular, there has not been extensive use of inferential techniques to learn about the underlying stochastic process that leads to failure via various phases of failure and repair.

Looking at reliability theory more generally, there has been a lot of work using Bayesian networks to model the dependencies present in systems (an overview is provided by Langseth and Portinale (2007)). Again, these have traditionally used common failure time distributions such as the Weibull and exponential. This has been limiting since the arcs in a Bayesian network are often used to represent 'causality', so there is no natural graph structure to explicitly represent a repairable redundant subsystem.

This paper is concerned with embedding phase-type distributions — which offer a natural expression of redundancy — into a Bayesian network. Thus, by effectively embedding an entire Markov process into a single node, it is possible to model the whole system and its dependencies.

In particular, the stochastic process underlying the phase-type distribution is carefully constructed to have a physical meaning, as opposed to the common practice of using general or Coxian phase-types simply as distribution fitting tools. As a result, inference on the Bayesian network can lead to explicit inference on failure and repair rates of the 'hidden' components of a redundant subsystem. The Bayesian approach, taken with mildly informative priors, is particularly crucial here due to the heavy overparametrization and hence non-identifiability of general phase-type distributions.

In Section 2 we review phase-type distributions and Bayesian networks describing their use in the context of this paper; in Section 3 we demonstrate simulation of a very simple example system expressed as a Bayesian network; in Section 4 we discuss inferential issues for the aforementioned model; in Section 5 we show some preliminary results; and in Section 6 discuss future work.

2 Background

2.1 Phase-type Distributions

Consider a continuous-time Markov chain on a finite discrete state space E with $n+1$ states, one of these states being absorbing. With a possible reordering of states, the generator of the Markov chain can be expressed as:

$$\mathbf{T} = \begin{pmatrix} \mathbf{S} & \mathbf{s} \\ \mathbf{0} & 0 \end{pmatrix}$$

In the context of modelling reliability of an $N+K$ redundant subsystem where individual components have exponentially distributed lifetimes, we may consider the first n states of the chain to represent the possible permutations of $0, \ldots, K$ unrepaired failures. Then $\mathbf{S} = (S_{ij})$ expresses the transition rates from state i to state j for $i, j \in \{1, \ldots, n\}$: in other words, the transitions to additional non-service affecting failures or to individual repairs. State $n+1$, the absorbing state, encompasses all situations where there are more than K failures which have not been repaired and thus the system is unavailable. Here $\mathbf{s} = (s_1, \ldots, s_n)^{\mathrm{T}}$ is the vector of transition rates from state i to the absorbing state.

Thus in a reliability context, this Markov process captures a redundant subsystem which can be repaired, and of particular interest would be the distribution of the time to failure — that is, the time until the Markov process enters state $n+1$.

Definition 1. A **phase-type distribution** (PHT) is the distribution of the time to entering the absorbing state of a Markov process with generator $\mathbf{T}$ and vector of initial state probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$.

The distribution and density can be expressed through the matrix-exponential as:

$$X \sim \mathrm{PHT}(\boldsymbol{\pi}, \mathbf{T}) \implies \begin{cases} F_X(x) &= 1 - \boldsymbol{\pi} \exp\{x\mathbf{S}\}\mathbf{e} \\ f_X(x) &= \boldsymbol{\pi} \exp\{x\mathbf{S}\}\mathbf{s} \end{cases}$$

where $\mathbf{e}$ is a vector of 1's of the appropriate dimension and x is the time to absorption.

Continuous time Markov chains are covered in Grimmett and Stirzaker (2001). Phase-type distributions were introduced in Neuts (1975) and an excellent survey of much of the theory is presented by Asmussen (2000).

2.2 Bayesian Networks

The primary motivation for setting up a model in terms of a Bayesian network is to make explicit the conditional independencies so as to simplify the subsequent inference of unobserved parts of the model. This is achieved by combining a directed acyclic graph whose nodes are the random variables in the model with conditional probability distributions for each node. The parsimony is achieved because each conditional probability distribution need only be defined conditional upon those nodes which comprise its parents in the graph. Thus,

$$\mathbb{P}(X_1 \,|\, X_2, \ldots, X_n) \equiv \mathbb{P}(X_1 \,|\, \mathrm{Pa}(X_1)) \quad \text{where} \quad \mathrm{Pa}(X_1) \subset \{X_2, \ldots, X_n\}$$

In the context of reliability, the most simple scenario may take the binary random variable for system availability to have one parent for availability of each critical subsystem. Each of those subsystem nodes would then have parents representing the priors of the parameters of the failure distributions, which could in turn have hyperpriors to express adjustment due to external factors such as environment.

The general theory of Bayesian networks is treated in detail by Pearl (1988). A survey of Bayesian networks as used by the reliability community and specifically contrasted with fault trees is presented by Langseth and Portinale (2007). A recent example of their use is in Flood et al. (2009).

3 Simulation of Network Reliability

Consider first simulation of reliability data given a model constructed using the ideas presented above. To aid simplicity of exposition, a small network of 5 nodes is simulated. Each node is considered to be a draw from a population of units of the same type.

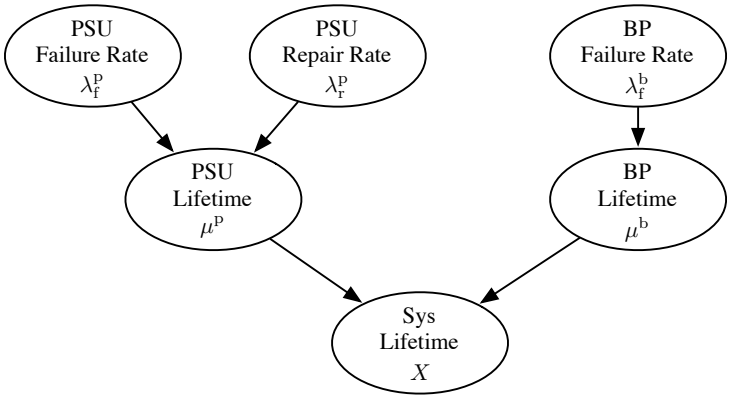

Figure 1: The Bayesian network (DAG part) for a network node.

3.1 The Network Nodes

Each node (a router, say) is considered to have two critical subsystems: a dual (1+1) hot-swappable redundant power supply (PSU) and a router backplane (BP). The lifetime of the redundant power supplies will be modelled as being of phase-type and the backplane as exponential. The lifetime of the unit is then the minimum of the lifetimes of the two subsystems.

This can be expressed in a Bayesian network as in Figure 1 with the following conditional probability distributions (with $k =$ shape, $\theta =$ scale for gamma variates and $\lambda =$ rate for exponential variates):

$$\begin{aligned}
\lambda_{\mathrm{f}}^{\mathrm{p}} &\sim \mathrm{Gamma}(k = 25, \theta = 1000000^{-1}) \\
\lambda_{\mathrm{r}}^{\mathrm{p}} &\sim \mathrm{Gamma}(k = 300, \theta = 10000^{-1}) \\
\lambda_{\mathrm{f}}^{\mathrm{b}} &\sim \mathrm{Gamma}(k = 20, \theta = 12000000^{-1}) \\
\\
\mu_{\mathrm{f}}^{\mathrm{p}} &\sim \mathrm{PHT}(\boldsymbol{\pi} = (1,0,0)^{\mathrm{T}}, \mathbf{T} = \mathbf{T}_{\mathrm{p}}) \\
\mu_{\mathrm{f}}^{\mathrm{b}} &\sim \mathrm{Exp}(\lambda = \lambda_{\mathrm{f}}^{\mathrm{b}}) \\
\\
X &= \min(\mu_{\mathrm{f}}^{\mathrm{p}}, \mu_{\mathrm{f}}^{\mathrm{b}})
\end{aligned}$$

where

$$\mathbf{T}_{\mathrm{p}} = \begin{pmatrix}
-2\lambda_{\mathrm{f}}^{\mathrm{p}} - 2\lambda_{\mathrm{f}}^{\mathrm{p}}/99 & \lambda_{\mathrm{f}}^{\mathrm{p}} & \lambda_{\mathrm{f}}^{\mathrm{p}} & 2\lambda_{\mathrm{f}}^{\mathrm{p}}/99 \\
\lambda_{\mathrm{r}}^{\mathrm{p}} & -\lambda_{\mathrm{r}}^{\mathrm{p}} - \lambda_{\mathrm{f}}^{\mathrm{p}} & 0 & \lambda_{\mathrm{f}}^{\mathrm{p}} \\
\lambda_{\mathrm{r}}^{\mathrm{p}} & 0 & -\lambda_{\mathrm{r}}^{\mathrm{p}} - \lambda_{\mathrm{f}}^{\mathrm{p}} & \lambda_{\mathrm{f}}^{\mathrm{p}} \\
0 & 0 & 0 & 0
\end{pmatrix}$$

Note that the entry $2\lambda_{\mathrm{f}}^{\mathrm{p}}/99$ was calculated to represent 1% uncovered failure rate. This follows from the fact that given a generator $\mathbf{G} = (g_{ij})$ for a continuous-time Markov chain (Grimmett and Stirzaker 2001),

$$g_{ii} := -\sum_{j:i\neq j} g_{ij} \qquad \text{and} \qquad \mathbb{P}(X_{n+1}(t+h) = j \,|\, X_n(t) = i \text{ and there } \textit{is} \text{ a transition after } h) = -\frac{g_{ij}}{g_{ii}}$$

thus

$$\mathbb{P}(j = 4 \,|\, i = 1, \text{transition after } h) = \frac{y}{2\lambda_{\mathrm{f}}^{\mathrm{p}} + y} = 0.01 \implies y = \frac{2\lambda_{\mathrm{f}}^{\mathrm{p}}}{99}$$

3.2 The Network

The elementary example network being modelled for availability is as seen in Figure 2. Each node X_i is a realisation from the Bayesian network model setup in Section 3.1. The network as a whole is considered to be 'available' if there exists a path from the entry arrow on the left to the exit arrow on the right through nodes that are themselves operational, and is otherwise considered to be in a 'failed' state.

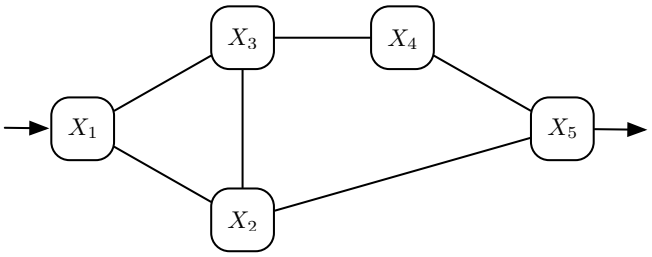

Figure 2: The network of nodes of the type outlined in Section 3.1.

Taking logical true to mean operational and logical false for failed, the availability of the whole network Y can be expressed as the logical operation:

$$Y = X_1 \wedge (X_2 \vee X_3) \wedge (X_2 \vee X_4) \wedge X_5$$

Under 100,000 repeated simulations performed in R (R Development Core Team 2009) this leads to the failure time kernel density estimate illustrated in Figure 3, with frequency of the four failure conditions as set out in Table 1. The kernel density estimate used is a standard non-parametric method for empirically estimating the probability density function of a random variable such that the estimate will be continuous and differentiable where this is true of the kernel function.

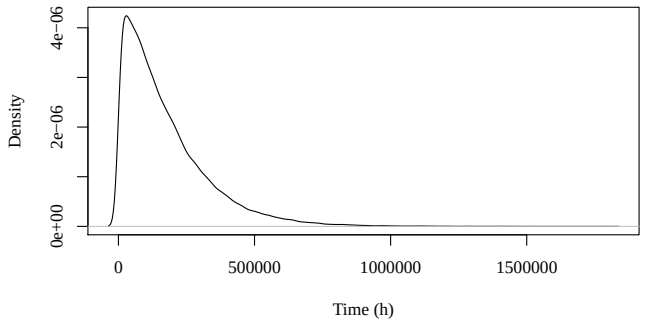

Figure 3: Empirical time to failure density for the network.

Table 1: Frequencies of failure states.

Failure	Freq.	Proportion
X_1	38120	38.1%
$X_2 \wedge X_3$	13326	13.3%
$X_2 \wedge X_4$	10091	10.1%
X_5	38463	38.5%

4 Bayesian Inference

The primary interest is inference on the parameters of the failure time distributions of nodes in a network. The key complexity is introduced by conducting inference on the phase-type distributional part of the model and so it

is this aspect which commands the remainder of this paper, since in principle it is straight forward to use standard techniques to perform inference on the Bayesian network once inference for the phase-type parts is available.

There has been effective theoretical work in the literature focusing on both frequentist (Asmussen et al. 1996) and Bayesian (Bladt et al. 2003) inferential procedures in a general setting.

The work by Asmussen et al. (1996) provides an implementation of the EM-algorithm for phase-type distributions and has seen very successful application in distribution fitting. However, general phase-type distributions are heavily over-parametrised leading to significant problems with identifiability of parameters. Indeed, a general phase-type of order p has $p^2 + p$ parameters but by a result in Section 2 of Asmussen et al. (1996), in fact the true dimension of the parameterisation is $2p - 1$. Thus there are many essentially equivalent parameter representations for any phase-type, making the EM-algorithm highly sensitive to local minima.

Bladt et al. (2003) provided a fully Bayesian framework for inference by constructing priors which were conjugate for the unobserved stochastic process. Because the ability to specify priors should enable injection of sufficient information to ensure that the posterior estimates reflect physical reality, rather than just one of many parameterisations for a distributional fit, this is the path chosen for this paper and is the work we seek to advance.

4.1 The Algorithm of Bladt et al.

The procedure developed in Bladt et al. (2003) is recapped in brief for the purpose of simplifying the elucidation of the innovations to be presented later. The algorithm is a Metropolis-Hastings within Gibbs Markov chain Monte Carlo sampler.

Consider observations $y_1, \ldots, y_N$. The approach reconstructs a sample from the underlying stochastic process, once for each observation via a Metropolis-Hastings step, with absorbing time equal to the observation time. Then summary statistics of this chain are used to update the parameters in a Gibbs step.

4.1.1 Algorithm 1: Sampling from the underlying Markov chain.

For each observation y_i from the phase-type there was an unobserved Markov chain (say $\mathbf{j}_i$) which traversed the state space until absorption at precisely time y_i. The chain $\mathbf{j}_i$ which led to the realisation of y_i was a sample from the distribution of all possible paths which absorb at time y_i: $\mathbb{P}(\text{path } \cdot \mid Y_i = y_i)$

Thus the goal is to sample a chain path $\mathbf{j}_i = (j_i(0), j_i(t_1), j_i(t_2), \ldots, j_i(y_i))$ where $j_i(y_i) = n + 1$ (the absorbing state) from the space of all such possible paths. This is achieved by instead sampling from the easier to simulate process where absorption occurs beyond time y_i.

1. Generate $\mathbf{j}_i = (j_i(0), j_i(t_1), \ldots, j_i(t_n))$ for $t_n < y_i$ from $\mathbb{P}(\text{path } \cdot \mid Y_i \geq y_i)$ by rejection sampling as the current step. In other words, store the chain truncated to the state just prior to the observed absorption time.
2. Generate $\mathbf{j}_i' = (j_i'(0), j_i'(t_1), \ldots, j_i'(t_n))$ for $t_n < y_i$ from $\mathbb{P}(\text{path } \cdot \mid Y_i \geq y_i)$ by rejection sampling as the proposal step.
3. Generate $U \sim \text{Unif}([0, 1])$
4. If $U \leq \min(1, s_{j_i(y_i^-)}/s_{j_i'(y_i^-)})$, then replace $\mathbf{j}_i$ with $\mathbf{j}_i'$ (replace current with proposal). The acceptance ratio is the ratio of the exit to absorption rates (from $\mathbf{s}$) of the states the chains were in just prior to the observation time.
5. Repeat from step 2 for a suitable burn-in time, then take only the last sample point and append $j_i(y_i) = n+1$ so that the chain absorbs at precisely y_i. For sufficient iterations, this will be a draw from $\mathbb{P}(\text{path } \cdot \mid Y_i = y_i)$.

Steps 1 and 2 are in fact easy to sample by simply using the generating matrix $\mathbf{T}$ of the continuous-time Markov chain and simulating from initial state through to absorption, rejecting until absorption occurs after time y_i. See Bladt et al. (2003) for further details.

4.1.2 Algorithm 2: Gibbs sampler for the phase-type distribution

It was shown that given the full Markov process, a Dirichlet prior on $\boldsymbol{\pi}$ and a gamma prior for each of the elements of $\mathbf{S}$ and $\mathbf{s}$ would provide conjugacy for the full conditionals of the posterior given the unobserved stochastic process. Thus Gibbs sampling can proceed, with the prior specification as follows:

$$\begin{aligned} s_i &\sim \text{Gamma}(k = \nu_{i0}, \theta = 1/\zeta_i) \\ S_{ij} &\sim \text{Gamma}(k = \nu_{ij}, \theta = 1/\zeta_i) \\ \boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)) \end{aligned}$$

Hence the full prior specification consists in selection of $\boldsymbol{\nu}, \boldsymbol{\zeta}$ and $\boldsymbol{\beta}$. In particular, note that the prior on all rates out of a state i must share common scale parameter $1/\zeta_i$.

The algorithm then proceeds as follows:

1. Generate $\boldsymbol{\pi}$, $S_{ij}(i \neq j)$ and s_i from the prior distribution, and set $S_{ii} = -s_i - \sum_{j:i\neq j} S_{ij}$
2. Use algorithm 1 to generate $(\mathbf{j}_1, \ldots, \mathbf{j}_n)$ where each $\mathbf{j}_i$ is a sample Markov process with absorption time y_i.
3. Calculate the combined statistics across all samples: $\mathbf{b} = \{B_i, i = 1, \ldots, p\}$, $\mathbf{z} = \{Z_i, i = 1, \ldots, p\}$, $\mathbf{N} = \{N_{ij}, i, j = 1, \ldots, p\}$ where B_i is the number of times a process started in state i, Z_i is the total time the process spent in state i, and N_{ij} is the total number of jumps $i \to j$.
4. Draw new parameter values from the conjugate full conditional distributions:

$$\begin{aligned} s_i &\sim \text{Gamma}(k = N_{i0} + \nu_{i0}, \theta = 1/(\zeta_i + z_i)) \\ S_{ij} &\sim \text{Gamma}(k = N_{ij} + \nu_{ij}, \theta = 1/(\zeta_i + z_i)) \\ \boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\beta} + \mathbf{b}) \end{aligned}$$

5. Repeat from step 2 for as many iterations as required.

See Bladt et al. (2003) for further details.

4.2 Revisions to the Algorithm

In the form presented above, the algorithm is excellent for performing general Bayesian inference on all parameters of a phase-type distribution in the situation where the conjugate priors are appropriate for conveying prior belief.

However, there are three issues with regard to this application:

- Certain state transitions make no physical sense and so their entry must be fixed at zero in the generator of the Markov chain. For example, from Section 3.1 a transition $2 \to 3$ is impossible since we cannot go from PS1 failed and PS2 operating, straight to PS1 operating and PS2 failed, without first going through a repair cycle for PS1.
- When performing inference on data where the phase-type distribution in question represents one of several (> 1) critical subsystems of a system, there will be some instances of system failure which were not a result of failure of the phase-type subsystem. However, the fact that the subsystem did not fail up to the time of the system failing *does* contain information: we have a right-censored observation. This is the phenomenon of competing risks and should also be taken into account (Prentice et al. 1978).
- Certain state transitions should (in an idealised sense) have identical parameters. For example, states $2 \to 1$ and $3 \to 1$ should have the same rate as there is no reason to believe a distributional difference in repairs to different power supplies. Moreover, this reasonable assumption will allow a significant reduction in the dimension of the parameter space.

Here we offer innovations to the algorithm of Bladt et al. (2003) so as to address such issues.

For the first, where entries should be constrained to zero, the parameter is fixed at zero when simulating step 2 and one does not draw new values in step 4. This solves the issue since we are effectively ceasing to consider this entry of $\mathbf{T}$ as a parameter and exclude it from all inferential procedures.

The second issue of right-censored observations can also be dealt with quite elegantly with a small change to the algorithm. The data will now consist of some actual and some censored observations: $\mathbf{y} = \{y_1, \ldots, y_m, y_{m+1}^{\text{c}}, \ldots, y_n^{\text{c}}\}$ where a superscript 'c' denotes right-censoring. Now when algorithm 1 is invoked by algorithm 2 it should be used as normal for $\{y_1, \ldots, y_m\}$, but for $\{y_{m+1}^{\text{c}}, \ldots, y_n^{\text{c}}\}$ it should instead only perform step 1 of algorithm 1 and return this value immediately as we have then sampled the underlying stochastic process conditional upon survival to at least $(y_i^{\text{c}})_{i=m+1}^n$. This allows us to effectively incorporate the fact that the unit survived to this time in the likelihood through the Metropolis-Hastings step.

The final issue requires examination of the posterior distribution which, when using the conjugate priors above and with the full data (full Markov processes, simulated by Metropolis-Hastings here), can be written (Bladt et al.

2003):

$$
\begin{aligned}
p(\boldsymbol{\pi}, \mathbf{T} \,|\, \boldsymbol{x}) \quad &\propto \quad \phi(\boldsymbol{\pi}, \mathbf{T}) p(\mathbf{x} \,|\, \boldsymbol{\pi}, \mathbf{T}) \\
&= \quad \left(\prod_{i=1}^{p} \pi_i^{\beta_i - 1} \prod_{i=1}^{p} s_i^{\nu_{i0}-1} e^{-s_i \zeta_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{\nu_{ij}-1} e^{-S_{ij}\zeta_i} \right) \left(\prod_{i=1}^{p} \pi_i^{B_i} \prod_{i=1}^{p} s_i^{N_{i0}} e^{-s_i Z_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{N_{ij}} e^{-S_{ij} Z_i} \right) \\
&= \quad \prod_{i=1}^{p} \pi_i^{B_i + \beta_i - 1} \prod_{i=1}^{p} s_i^{N_{i0} + \nu_{i0} - 1} e^{-s_i(\zeta_i + Z_i)} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{N_{ij} + \nu_{ij} - 1} e^{-S_{ij}(Z_i + \zeta_i)}
\end{aligned}
$$

Thus, where we constrain parameters in different parts of the generator $\mathbf{S}$ to be equal, the resultant full conditional posteriors still keep the property of conjugacy. For example, if we want to constrain $\mathcal{C} : S_{12} = S_{13} = s_2 = s_3 = \lambda_{\mathrm{f}}^{\mathrm{p}}$ (as for the model outlines in Section 3.1), then the full conditional posterior for $\lambda_{\mathrm{f}}^{\mathrm{p}}$ becomes:

$$
\begin{aligned}
p(\lambda_{\mathrm{f}}^{\mathrm{p}} \,|\, \mathbf{x}, \mathcal{C}) \quad &\propto \quad (\lambda_{\mathrm{f}}^{\mathrm{p}})^{N_{12}+N_{13}+N_{20}+N_{30}+\nu-1} e^{\lambda_{\mathrm{f}}^{\mathrm{p}}(2Z_1+Z_2+Z_3+\zeta)} \\
&\sim \quad \mathrm{Gamma}(k = N_{12} + N_{13} + N_{20} + N_{30} + \nu, \theta = 1/(2Z_1 + Z_2 + Z_3 + \zeta))
\end{aligned}
$$

where ν and ζ were the hyperprior parameters for $\lambda_{\mathrm{f}}^{\mathrm{p}}$. This works likewise for other constraints for parameter equality and reduces the dimensionality of the parameter space significantly.

5 Results

The theory from Section 4 was applied to perform the difficult part of the inference on a network of the type illustrated in Section 2: inferring parameters of a phase-type distribution. The results here are on simulated data for 'toy example' parameter values.

5.1 MCMC Runs

The modified algorithm 2 was run on simulated data for both uncensored and censored cases. Two chains of 20,000 iterations each were produced for each run.

Then, for each run autocorrelation plots were used to determine the level of thinning required (approximately lag 20 in most cases).

After this, three tests of stationarity were applied:

- Gelman-Rubin Shrink Factor (Gelman and Rubin 1992): intuitively, if behaviour is good and stationarity has been achieved, then all chains should behave as a single homogeneous group. Thus, the Gelman-Rubin shrink factor effectively performs an ANOVA between chains. Brooks and Gelman (1998) indicate a shrink factor of under 1.2 to show approximate convergence, with under 1.1 being a more stringent option.

- Heidelberger and Welch's Convergence Diagnostic (Heidelberger and Welch 1983): which uses the Cramér-von-Mises statistic to empirically estimate convergence to stationarity. The test is either pass/fail for a given burn-in based on the critical values of the Cramér-von-Mises test statistic.

- The Geweke statistic (Geweke 1991): where two non-overlapping parts of a single chain are taken (a standard in the literature is the first 10% and last 50% for technical reasons of power and independence) and a simple difference of means test based on Geweke's statistic is applied to the two samples. This produces a Z-score which can be used to test if there is any significant difference.

All the above tests can be applied with progressively increasing burn-in sizes to gain insight into the necessary burn-in to reach stationarity. Each run was independently subjected to these tests and the largest of the three recommended burn-ins applied. This averaged 50-100 iterations after thinning (1000-2000 equivalent pre-thinning) over the different runs.

5.2 Uncensored Phase-Type Inferential Results

First to be considered is the uncensored case. Data were simulated from a phase-type distribution with parameters:

$$\boldsymbol{\pi} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{S} = \begin{pmatrix} -3.6 & 1.8 & 1.8 \\ 9.5 & -11.3 & 0 \\ 9.5 & 0 & -11.3 \end{pmatrix}$$

In other words, the distribution models a dual redundant system with failure rate (λ_{f}) 1.8, repair rate (λ_{r}) 9.5 and no uncovered failures. Priors were specified which meant these true values were away from the prior mode, the full specification being:

$$\lambda_{\mathrm{f}} \sim \mathrm{Gamma}(k = 24, \theta = 1/16) \qquad \text{and} \qquad \lambda_{\mathrm{r}} \sim \mathrm{Gamma}(k = 180, \theta = 1/16)$$

Initially, 100 failure times were used to perform inference. This resulted in posterior plots shown in Figure 4, which demonstrate very successful results for the failure rate and no change from the prior for the repair rate. This strongly indicates that 100 data points is insufficient to convey information about the repair rate, probably due to the fact so little time is spend in a degraded state. The 95% highest posterior density intervals are shown in Table 2

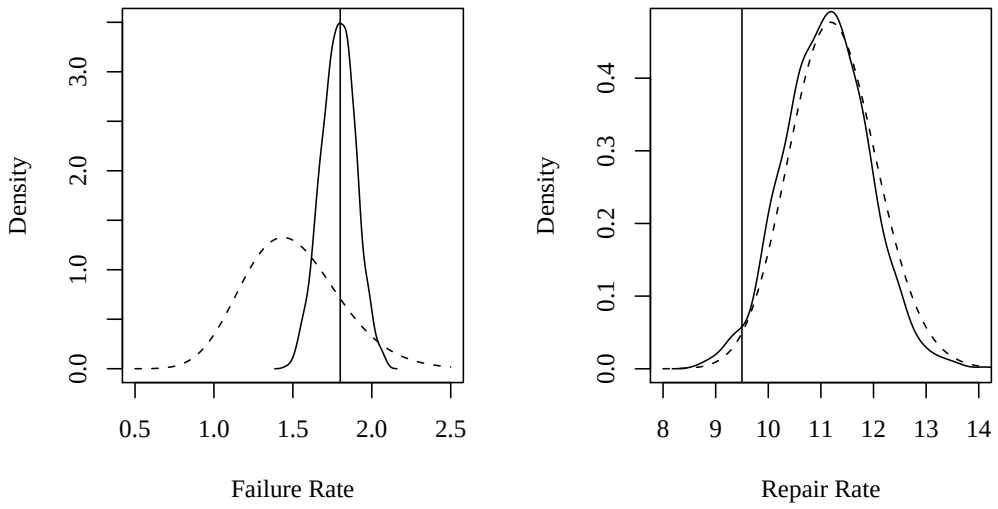

Figure 4: Uncensored data posterior densities in solid line; prior in dashed line; true values at the vertical line.

Table 2: 95% highest posterior density intervals.

Parameter	Lower	Upper
Failure Rate	1.58	2.00
Repair Rate	9.60	12.67

5.3 Censored Phase-Type Inferential Results

Again, simulations were used to generate data upon which the inferential procedure could be tested using the same true values for λ_{f} and λ_{r} as in the uncensored case. This phase-type was then put in competing risk with

an exponential random failure variate with rate ≈ 0.144, leading to around 25% of the data being right-censored. Again, to facilitate direct comparability the same priors as the uncensored case were employed.

The resultant posterior plots can be seen in Figure 5. Interestingly, the censored case seemed to carry a little more information about the repair time, though at the expense of less accurate modal estimation of the failure rate. More data points would likely be required to observe an improvement in the tightness of these posterior plots.

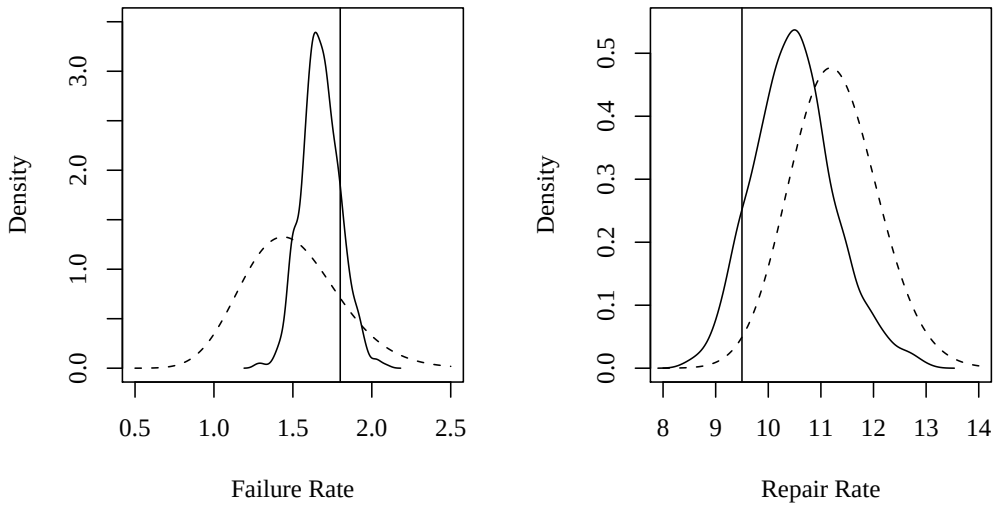

Figure 5: Censored data posterior densities in solid line; prior in dashed line; true values at the vertical line.

6 Conclusions & Future Work

Innovations to the algorithm of Bladt et al. (2003) have been presented, which extend this general purpose inferential procedure into one that enables a significant reduction in parameter dimensionality. This, combined with prior information, enables scientific meaning to be placed on the parameter values where the underlying Markov process has a physical interpretation. Furthermore, the techniques developed enable the use of censored data.

The improved algorithm performed well when applied to a 'toy example', showing that the posterior distribution of the failure rate successfully captures information about the true parameter value. More data is necessary to infer anything about repair rates due to the relatively short period of lifetime spend in a degraded state. Indeed, initial work indicates that adding a significant number of observations starts to give accurate posterior estimates for both parameters in the example presented.

There is significant scope for future work. Inference on the coverage rate for uncovered failures was not presented here and sensible adjustments to the algorithm may be required to provide scientifically interpretable results.

A particular aspect the authors intend to focus on is improvements to the speed of MCMC runs. At present, for more realistic parameter values the computational overhead reaches unacceptable levels. In particular, algorithm 1 has been identified as the source of slowdown. Preliminary work has already commenced on modifications which it is hoped will significantly reduce the dependance of the runtime on parameter values.

Finally, although it is now straightforward in principle to incorporate this inferential procedure into the Bayesian network setup of Section 2, a deeper analysis of so doing is called for once speed is improved.

Appendix

Computational Notes

The code was written in R (R Development Core Team 2009) and steps 1 and 2 of algorithm 1 were written natively in C to make computation times tractable. Additionally, step 2 (*ie* all of algorithm 1) was multi-threaded to provide additional speedup, since computation of each $\mathbf{j}_i$ is entirely independent. Finally, it is noted that for algorithm 2 it is not in fact necessary to store the entire chains, merely track $\mathbf{b}, \mathbf{z}$ and $\mathbf{N}$ for each chain which can then be summed over chains in step 3 of algorithm 2.

In the case of inference on 100 observations as presented in Section 5, a runtime of about 34 minutes produced 20,000 simulations on a dual core 2.4GHz Intel® Core™ 2 Duo processor.

Thus, for failure and repair rates which are not too many orders of magnitude different, MCMC run times are reasonable. However, where the two parameters are large orders of magnitude apart, iterations of algorithm 2 take an impracticably long time to execute.

Acknowledgements

This work was funded by the Irish Research Council for Science, Engineering and Technology (IRCSET). The authors would like to thank Brett Houlding of Trinity College Dublin for his helpful comments.

References

Asmussen, S. (2000). Matrix-analytic models and their analysis. *Scandinavian Journal of Statistics 27*(2), 193–226.

Asmussen, S., O. Nerman, and M. Olsson (1996). Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics 23*(4), 419–441.

Bladt, M., A. Gonzalez, and S. L. Lauritzen (2003). The estimation of phase-type related functionals using Markov chain Monte Carlo methods. *Scandinavian Journal of Statistics 2003*(4), 280–300.

Brooks, S. P. and A. Gelman (1998, Dec). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics 7*(4), 434–455.

Flood, B., B. Houlding, S. P. Wilson, and S. Vilkomir (2009). A probability model of system downtime with implications for optimal warranty design. *Quality and Reliability Engineering International*.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*(4), 457–511.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*.

Grimmett, G. R. and D. R. Stirzaker (2001). *Probability and Random Processes* (3rd ed.). Oxford University Press.

Heidelberger, P. and P. D. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research 31*(6), 1109–1144.

Langseth, H. and L. Portinale (2007). Bayesian networks in reliability. *Reliability Engineering & System Safety 92*(1), 92–108.

Neuts, M., R. Pérez-Ocón, and I. Torres-Castro (2000). Repairable models with operating and repair times governed by phase type distributions. *Advances in Applied Probability 32*(2), 468–479.

Neuts, M. F. (1975). Probability distributions of phase type. *Liber Amicorum Prof. Emeritus H. Florin, Dept. Math, Univ. Louvain, Belgium*, 173–206.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Prentice, R. L., J. D. Kalbfleisch, J. A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow (1978). The analysis of exponentially distributed life-times with two types. *Biometrics 34*(4), 541–554.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.