A Comparison of two Sequential Metropolis-Hastings Algorithms with Standard Simulation Techniques in Bayesian Inference in Reliability Models Involving the Generalized Gamma Distribution.

Jørund Gåsemyr (gaasemyr@math.uio.no)
Department of Mathematics, University of Oslo,
Bent Natvig (bent@math.uio.no)
Department of Mathematics, University of Oslo
and Norwegian Computing Center,
Erik Sørensen (erik.k.m.sorensen@PMSC.no)
Policy Management Systems Corporation, Oslo

Abstract. In this paper we consider the generalized gamma distribution as introduced in Gåsemyr and Natvig (1998). This distribution enters naturally in Bayesian inference in exponential survival models with left censoring. In the paper mentioned above it is shown that the weighted sum of products of generalized gamma distributions is a conjugate prior for the parameters of component lifetimes, having autopsy data in a Marshall-Olkin shock model. A corresponding result is shown in Gåsemyr and Natvig (1999) for independent, exponentially distributed component lifetimes in a model with partial monitoring of components with applications to preventive system maintenance. A discussion in the present paper strongly indicates that expressing the posterior distribution in terms of the generalized gamma distribution is computationally efficient compared to using the ordinary gamma distribution in such models. Furthermore, we present two types of sequential Metropolis-Hastings algorithms that may be used in Bayesian inference in situations where exact methods are intractable. Finally these types of algorithms are compared with standard simulation techniques and

analytical results in arriving at the posterior distribution of the parameters of component lifetimes in special cases of the mentioned models. It seems that one of these types of algorithms may be very favourable when prior assessments are updated by several data sets and when there are significant discrepancies between the prior assessments and the data.

Key words: Exponential survival models, left censoring, autopsy data, Marshall-Olkin shock model, preventive system maintenance.

1 Bayesian inference in reliability models involving the generalized gamma distribution

In this paper we consider the generalized gamma distribution, as introduced in Gåsemyr and Natvig (1998), given by the following definition.

Definition 1.1 *For positive real numbers $a, b, t_1, \ldots, t_m$, $m \geq 0$, define the functions*

$$h(\theta; a, b, \boldsymbol{t}) = \theta^{a-1} \exp(-b\theta) \prod_{i=1}^{m} (1 - \exp(-\theta t_i)) \ , \qquad \theta \geq 0 \ , \tag{1.1}$$

$$h(\theta; a, b) = \theta^{a-1} exp(-b\theta), \theta \geq 0,$$

where $\boldsymbol{t} = (t_1, \ldots, t_m)$. Define the normalizing constant $\gamma(a, b, \boldsymbol{t})$ by

$$(\gamma(a, b, \boldsymbol{t}))^{-1} = \Gamma(a) \sum_{\boldsymbol{d} \in \{0,1\}^m} (-1)^{|\boldsymbol{d}|} (b + \boldsymbol{d} \cdot \boldsymbol{t})^{-a} \ , \tag{1.2}$$

where $|\boldsymbol{d}| = d_1 + \cdots + d_m$. The generalized gamma distribution with parameters a, b and $\boldsymbol{t}$ is then defined as the probability distribution on $[0, \infty)$ with density function given by

$$g(\theta; a, b, \boldsymbol{t}) = \gamma(a, b, \boldsymbol{t}) h(\theta; a, b, \boldsymbol{t}) \ , \qquad \theta \geq 0 \tag{1.3}$$

The ordinary gamma distribution, $g(\theta; a, b)$, is the special case corresponding to $m = 0$.

This distribution enters naturally as the posterior distributions for failure rates in exponential survival models with left censoring as seen in the following.

Example 1.1 Suppose r identical components are put on test in separate test chambers under identical conditions. Under the experimental conditions, the components have independent exponential life distributions with failure rate θ. The ith component is inspected at a deterministic inspection time t_i, and it is observed whether the component has failed before t_i. The inspection interfers with the experimental conditions, so further testing of the component cannot be done. The inspection may itself for instance be destructive to the component, or its failure rate may be increased due to stress related to the inspection. À priori θ is assumed to be gamma distributed with parameters a, b. Let T_i be the lifetime of the ith component, interpreted as the potential lifetime that would have resulted if the experiment had not been interrupted. Define $D_i = I(T_i > t_i)$. The likelihood function for the data $D_i = d_i$, $i = 1, \ldots, r$, is then

$$L(\theta|d_1, \ldots, d_r) = \prod_{i=1}^{r} P(D_i = d_i|\theta) = \prod_{i=1}^{r} (\exp(-\theta t_i))^{d_i} (1 - \exp(-\theta t_i))^{1-d_i}$$

If $d_i = 0$ for $i = i_1, \ldots, i_m$, $d_i = 1$ otherwise, we obtain by Bayes theorem the posterior distribution

$$\pi(\theta|d_1, \ldots, d_r) = g(\theta; a, b + \boldsymbol{d} \cdot \boldsymbol{t}, t_{i_1}, \ldots, t_{i_m})$$

In order to compute the normalizing constant of this density analytically, we see from (1.2) and (1.3) that we must add up 2^m terms. If m, the number of left censored lifetimes, is large, the computational complexity can be an obstacle and simulation is the only alternative. Furthermore, this density function can be approximated arbitrarily well if we can generate a sufficiently large sample from the posterior distribution. To derive the density function, the sample is used as input in a standard density estimate. It is easy to extend the results above to a situation where the ith component is continuously monitored in the interval (s_i, t_i), $i = 1, \ldots, n$.

We now consider a binary, monotone system (E, ϕ), where $E = \{1, \ldots, n\}$ is the set of components and ϕ is the structure function describing the state of the system in terms of the binary states of the components. Denote the lifetime of the system by T and the lifetime of the ith component by T_i. The state of the ith component at time t is denoted $X_i(t)$ and we have $X_i(t) = I(T_i > t)$, $i \in E$. Let $\boldsymbol{X}(t) = (X_1(t), \ldots, X_n(t))$. We then have $\phi(\boldsymbol{X}(t)) = I(T > t)$. The autopsy data of the system is the pair (T, D), where $D = \{i|T_i \leq T\}$, the set of failed components by the time of system failure, see Meilijson (1981).

Following Gåsemyr and Natvig (1998) consider a set of shocks $S = \{1, 2, \ldots, n+p\}$. Here, $1, 2, \ldots, n$ represent individual shocks destroying the corresponding components of E, whereas $n+1, \ldots, n+p$ represent common shocks; the lth shock destroying the components in $D_l \subset E$. Let V_l be the time until the lth shock occurs. Thus, $T_i = min\{V_l | i \in D_l\}$, $i = 1, \ldots, n$. Now assume that $V_1, \ldots, V_{n+p}$ are independent, exponentially distributed with failure rates $\theta_1, \ldots, \theta_{n+p}$; i.e. we have a Marshall-Olkin shock model, see Marshall and Olkin (1967). A subset $A \subset E$ satisfies $P(D = A) > 0$ iff A is a cut set, i.e. the system has failed if all components in A have failed, and there exists a shock $l \in S$ such that $D_l \subset A$ and $A - D_l$ is not a cut set. Such a shock l is called a critical shock and such a subset A a fatal set. Denote the set of critical shocks for A by C_A and let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be the set of fatal sets. Define $G_A(t) = P(T \leq t, D = A)$ with density function $g_A(t) = \frac{d}{dt} G_A(t)$. Gåsemyr and Natvig (1998) gives a computationally efficient procedure for deriving this likelihood, leading to the following lemma.

Lemma 1.2 *For the Marshall-Olkin shock model with failure rates $\theta_1, \ldots, \theta_{n+p}$ the likelihood function can be written in the form:*

$$L(\boldsymbol{\theta}) = \sum_{k=1}^{K} \prod_{l \in B_k} h(\theta_l; 1, 0, t) \prod_{l \in C_k} h(\theta_l; 1, t) h(\theta_{j_k}; 2, t), \tag{1.4}$$

where B_k, C_k, $\{j_k\}$ are disjoint subsets of S for each $k = 1, \ldots, K$.

This leads almost immediately to the following main result in Gåsemyr and Natvig (1998) on Bayesian inference based on autopsy data.

Theorem 1.3 *a) Suppose that the failure rates θ_l, $l = 1, \ldots, n+p$ for the Marshall-Olkin shock model for a binary, monotone system (E, ϕ) have a joint prior distribution of the form*

$$\begin{aligned} \pi_0(\boldsymbol{\theta}) &\propto \sum_{j=1}^{J} \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l}) \\ &= \sum_{j=1}^{J} \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l})^{-1} g(\theta_l; a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l}) \end{aligned} \tag{1.5}$$

Then the posterior distribution of $\boldsymbol{\theta}$ given the autopsy data $(T = t, D = A)$ with likelihood function given by (1.4) is of the form

$$\begin{aligned} \pi(\boldsymbol{\theta} | t, A) \propto \sum_{j=1}^{J} \sum_{k=1}^{K} \{ &\prod_{l \in B_k} h(\theta_l; a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l}, t) \prod_{l \in C_k} h(\theta_l; a_{j,l}, b_{j,l} + t, \boldsymbol{t}_{j,l}) \} \\ &\times h(\theta_{j_k}; a_{j,j_k} + 1, b_{j,j_k} + t, \boldsymbol{t}_{j,j_k}) \prod_{l \in S - (B_k \cup C_k \cup j_k)} h(\theta_l; a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l}) \end{aligned} \tag{1.6}$$

b) The class of distributions of the form (1.5) is a conjugate class of prior distributions for the exponential autopsy shock model.

c) Suppose the prior distribution for $\boldsymbol{\theta}$ *is given by:*

$$\pi_0(\boldsymbol{\theta}) = \prod_{l=1}^{n+p} g(\theta_l; a_l, b_l) \tag{1.7}$$

and that it is updated with autopsy data from r *independent systems, of the type* $(T_1 = t_1, D_1 = A_{i_1}, \ldots, T_r = t_r, D_r = A_{i_r})$. *Then the posterior distribution is of the form (1.5) with* $\boldsymbol{t}_{j,l}$ *a subvector of* $\boldsymbol{t} = (t_1, t_2, \ldots, t_r)$ *for all* $j = 1, \ldots, J$, $l = 1, \ldots, n+p$.

Theorem 1.3 b) states that the weighted sum of products of generalized gamma distributions is a conjugate prior for $\boldsymbol{\theta}$ in the Marshall-Olkin shock model.

In Gåsemyr and Natvig (1999) a monitoring scheme is considered in which a subset $M = \{1, \ldots, p\}$ of components with independent lifetimes is monitored from time 0 onwards, while the components in another subset $C = \{p+1, \ldots, p+q\}$, where $1 \leq p < p+q \leq n$, are conditionally monitored, i.e. they are monitored from certain time points τ_i onwards, $i \in C$. These time points are called inspection times and are determined by the observed history of the system according to a specific strategy determined in advance. For any component i for which $T_i \leq T$, the failure time is recorded if it is subject to monitoring at that time. In addition to data arising from this monitoring scheme, autopsy data are observed, if not censored.

For this model Gåsemyr and Natvig (1999) arrives at the following lemma, being very similar to Lemma 1.2.

Lemma 1.4 *For the case of exponentially distributed component lifetimes with failure rates* $\theta_1, \ldots, \theta_n$ *the likelihood function of Theorem* 2.1 *of Gåsemyr and Natvig (1999) can be written in the form*

$$L(\boldsymbol{\theta}) = \sum_{k=1}^{K} \prod_{l \in B_k} h(\theta_l; 1, 0, t_{k,l}) \prod_{l \in C_k} h(\theta_l; 1, t_{k,l}) \prod_{l \in D_k} h(\theta_l; 2, t_{k,l}) \,, \tag{1.8}$$

where B_k, C_k, D_k *are disjoint subsets of* E *for each* $k = 1, \ldots, K$.

This leads to Theorem 5.3 of Gåsemyr and Natvig (1999) being an obvious modification of Theorem 1.3 above.

A wider class of priors is obtained if one allows for a positive weight w_j for the jth summand in (1.5). An even wider class of conjugate priors is

the class of distributions of the form

$$\pi_0(\boldsymbol{\theta}) = \sum_{j=1}^{J} w_j \prod_{l=1}^{n+p} g(\theta_l; a_{j,l}, b_{j,l}) \ , \tag{1.9}$$

where in this case, we have to allow for negative weights w_j. The requirement is that $\sum_{j=1}^{J} w_j = 1$, and that $\pi_0(\boldsymbol{\theta}) \geq 0$ for all vectors $\boldsymbol{\theta}$ with positive entries. Posterior densities of the form (1.9) arise, if the factors $(1 - exp(-\theta_l t))$, $l \in B_k$, $k = 1, \ldots, K$ appearing in (1.4) (and correspondingly in (1.8)) are multiplied out. However, as the discussion in Appendix 1 strongly indicates, expressing the posterior distribution in terms of the generalized gamma distribution as in (1.5) is computationally efficient compared to using the ordinary gamma distribution as in (1.9).

In any case the form (1.5) has considerable conceptual advantages. It expresses the density in terms of functions, all of whose parameters are easily interpretable; i.e. $a_{j,l}$, $b_{j,l}$ and $\boldsymbol{t}_{j,l}$ represent respectively the number of failed components whose exact failure times are known, the total time on test for the components that are not left censored, and the censoring times for the components that are left censored. In the form (1.9), interpretation of the parameters is much more difficult (see (A1.2) of Appendix 1) especially since the sum contains both positive and negative terms.

Often one is interested in estimating $\boldsymbol{\theta}$. The standard Bayes estimate, minimizing the expected quadratic loss, is for the Marshall-Olkin shock model

$$E(\theta_i) = \frac{\sum_{j=1}^{J} \prod_{l=1}^{n+p} \gamma(a_{j,l} + I(l=i), b_{j,l}, \boldsymbol{t}_{j,l})^{-1}}{\sum_{j=1}^{J} \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l})^{-1}} \ , \quad i = 1, \ldots, n+p \tag{1.10}$$

Another quantity of interest is the predictive system survival probability, which for the Marshall-Olkin shock model is given by

$$P(T > t) = \frac{\sum_{A \subset E} \sum_{j=1}^{J} \delta_A \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l} + I(l \in E_A)t, \boldsymbol{t}_{j,l})^{-1}}{\sum_{j=1}^{J} \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l})^{-1}} \ , \tag{1.11}$$

where δ_A, $A \subset E$, is the signed domination function of (E, ϕ) defined through

$$\phi(\boldsymbol{x}) = \sum_{A \subset E} \delta_A \prod_{k \in A} x_k$$

The proofs of (1.10) and (1.11) are straightforward from (1.5) and given in Gåsemyr and Natvig (1998). For the model of Gåsemyr and Natvig (1999) completely parallel results hold.

In Section 2 we present two types of sequential Metropolis-Hastings algorithms that may be used in Bayesian inference in situations where exact

methods are intractable. These types of algorithms are in Sections 3 and 4 compared with standard simulation techniques and analytical results in arriving at the posterior distribution of $\boldsymbol{\theta}$ in special cases of the models respectively in Gåsemyr and Natvig (1999) and Gåsemyr and Natvig (1998). In Section 5 some concluding remarks are given. It seems that one of these types of algorithms may be very favourable when prior assessments are updated by several data sets and when there are significant discrepancies between the prior assessments and the data.

2 Two sequential Metropolis-Hastings algorithms

As already mentioned at the end of Example 1.1 simulation may be the only alternative to compute the generalized gamma density for large m. We now return to the expressions for $E(\theta_i)$, $i = 1, \ldots, n+p$ and $P(T > t)$ given respectively by (1.10) and (1.11) for the Marshall-Olkin shock model. If the right hand side of (1.5), entering in (1.10) and (1.11), results from updating of a prior distribution π_0 with independent autopsy data from r systems, the potential number of summands involved in the computation of each of the normalizing constants $\gamma(a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l})^{-1}$, see (A1.1), increases with a factor of 2 for each new observation, i.e. with a factor of 2^r altogether. Furthermore, the number of summands appearing in (1.10) and (1.11) may increase drastically as r increases, see Theorem 1.3 a). For instance, this number increases with a factor of $K = 5$ for each new observation in the simple example in Section 4. Thus the computational complexity may be formidable. In some cases it may even be impossible to calculate the signed domination function δ in (1.11). It may also be of interest to calculate the marginal posterior density for each θ_i, $i = 1, \ldots, n+p$. This requires integrating out the other θ_i's from the joint posterior density, an operation which makes the computational task even more challenging. To overcome such problems simulation can be the only alternative.

We will now present two types of sequential Metropolis-Hastings algorithms. Basic papers in this area are Smith and Roberts (1993), Tierney (1994), Besag et al. (1995) and Chib and Greenberg (1995) with corresponding discussions and references. Let us begin with the ordinary Metropolis-Hastings algorithm and assume it is easy to simulate from the prior distribution π_0, which for instance can be given by (1.7). Following Gåsemyr and Natvig (1998) we denote by $L(\boldsymbol{\theta}|t_i, A_{j_i})$ the likelihood for $\boldsymbol{\theta}$ given the

autopsy data ($T_i = t_i, D_i = A_{j_i}$). The posterior distribution for $\boldsymbol{\theta}$, given autopsy data from r independent systems, is hence

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|t_1, A_{j_1}, \ldots, t_r, A_{j_r}) \propto \pi_0(\boldsymbol{\theta}) \prod_{i=1}^{r} L(\boldsymbol{\theta}|t_i, A_{j_i}) \tag{2.1}$$

We want to simulate a Markov chain $\{\boldsymbol{\theta}_k\}$ whose stationary distribution is π. Actually, in the terminology of Tierney (1994), we suggest simulating from an independent chain with fixed proposal density π_0. We call this algorithm Parametric Independent Chain (PIC). We then start with an arbitrary $\boldsymbol{\theta}$, e.g. a value drawn from π_0. Given $\boldsymbol{\theta}_k$, draw $\boldsymbol{\theta}'$ from π_0. Put $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}'$ with acceptance probability $\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}') = \min\{1, \beta(\boldsymbol{\theta}_k, \boldsymbol{\theta}')\}$, where

$$\begin{aligned} \beta(\boldsymbol{\theta}_k, \boldsymbol{\theta}') &= (\pi(\boldsymbol{\theta}')\pi_0(\boldsymbol{\theta}_k))/(\pi(\boldsymbol{\theta}_k)\pi_0(\boldsymbol{\theta}')) \\ &= (\prod_{i=1}^{r} L(\boldsymbol{\theta}'|t_i, A_{j_i}))/(\prod_{i=1}^{r} L(\boldsymbol{\theta}_k|t_i, A_{j_i})) \end{aligned} \tag{2.2}$$

With probability $1 - \alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}')$ we put $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$. The predictive survival probability, $P(T > t)$, for the system may then be estimated by

$$N^{-1} \sum_{k=1}^{N} P(T > t|\boldsymbol{\theta}_k) \ , \tag{2.3}$$

where N is chosen sufficiently large to ensure convergence, possibly after a burn-in period. If the exact reliability of the system is hard to calculate, an approximation to (2.3) may be obtained by replacing the summands of (2.3) by approximate values based for instance on the bounds for the reliability of a shock system given in Gåsemyr and Natvig (1995).

If autopsy data ($T_{r+1} = t_{r+1}, D_{r+1} = A_{j_{r+1}}$) from another system is obtained, one must in principle repeat the procedure. Note that it is reasonable to expect that the Markov chain converges faster the closer π_0 is to π. One would therefore expect convergence to $\pi(\boldsymbol{\theta}|t_1, A_{j_1}, \ldots, t_{r+1}, A_{j_{r+1}})$ to be faster if drawing candidate values $\boldsymbol{\theta}'$ from π_0 could be replaced by drawing from $\pi(\boldsymbol{\theta}|t_1, A_{j_1}, \ldots, t_r, A_{j_r})$. It may therefore be profitable to draw from an easily simulated approximation, $\pi^{(r)}(\boldsymbol{\theta})$, to the latter distribution. This is the idea which the two following types of sequential Metropolis-Hastings algorithms are based on. Furthermore, for both algorithms the proposal distribution, $\pi^{(r)}(\boldsymbol{\theta})$, is assumed to be a product of its marginal distributions.

In the first type of algorithm $\pi^{(r)}(\boldsymbol{\theta})$ is a product of gamma distributions with the correct marginal expectations and correct variances. We call this algorithm Parametric Sequential Independent Chain (PSIC). This seems to

be an original suggestion of an "adaptive sampler", the design of which is a legitimate goal according to Besag et al. (1995), see page 61. Note that $E(\theta_l|t_1, A_{j_1}, \ldots, t_r, A_{j_r})$ and $E(\theta_l^2|t_1, A_{j_1}, \ldots, t_r, A_{j_r})$, $l = 1, \ldots, n+p$, may be estimated by replacing $P(T > t|\boldsymbol{\theta}_k)$ with $\theta_{k,l}$ and $\theta_{k,l}^2$ respectively in (2.3). The original simulation procedure would then be modified by changing $\beta(\boldsymbol{\theta}_k, \boldsymbol{\theta}')$ to

$$\beta(\boldsymbol{\theta}_k, \boldsymbol{\theta}') = (\pi_0(\boldsymbol{\theta}') \prod_{i=1}^{r+1} L(\boldsymbol{\theta}'|t_i, A_{j_i})\pi^{(r)}(\boldsymbol{\theta}_k))/(\pi_0(\boldsymbol{\theta}_k) \prod_{i=1}^{r+1} L(\boldsymbol{\theta}_k|t_i, A_{j_i})\pi^{(r)}(\boldsymbol{\theta}')) \quad (2.4)$$

The effect on the convergence rate of choosing a prior and a distribution for candidate values $\boldsymbol{\theta}'$ as close as possible to the distribution that we want to simulate, may justify choosing more complicated distributions than products of gamma distributions; for example products of generalized gamma distributions or even convex combinations of such. Thus, generalized gamma distributions may play a role even when we have to resort to simulation.

The second type of sequential Metropolis-Hastings algorithm is based on a smoothing technique given in Bølviken, Christophersen and Storvik (1998), see page 128. We then draw S samples from $\pi(\boldsymbol{\theta}|t_1, A_{j_1}, \ldots, t_r, A_{j_r})$ and let $\pi^{(r)}(\boldsymbol{\theta})$ be a joint density which is the product of the smoothed marginal histograms. We call this algorithm Non Parametric Sequential Independent Chain (NPSIC). Denote the S samples by $\boldsymbol{\theta}^{*(r)}(s)$, $s = 1, \ldots, S$. We first derive a set of corrected samples each assigned a probability $1/S$.

$$\theta_i^{**(r)}(s) = \overline{\theta}_i^{*(r)} + [(1-\lambda^2)S/(S-1)]^{1/2}[\theta_i^{*(r)}(s) - \overline{\theta}_i^{*(r)}], \qquad s = 1, \ldots, S$$

Here $\overline{\theta}_i^{*(r)}$ is the mean of $\{\theta_i^{*(r)}(s)\}_{s=1}^S$. Let $\sigma_i^{2*(r)}$ be the corresponding sample variance. $\lambda \in [0, 1]$ is a user selected parameter. It is straightforward to verify that the distribution has mean $\overline{\theta}_i^{*(r)}$ and variance $(1 - \lambda^2)\sigma_i^{2*(r)}$. Now take the corrected sample $\theta_i^{**(r)}(s)$ and add an independent Gaussian variable $\epsilon_i^{**(r)}(s)$ with mean zero and variance $\lambda^2\sigma_i^{2*(r)}$. Clearly the resulting distribution has a mean and variance equal to $\overline{\theta}_i^{*(r)}$ and $\sigma_i^{2*(r)}$ respectively. This distribution is a continuous one and thus represents a smoothed version of the marginal histogram.

As presented above, the proposal distribution is updated once for each data set, the starting point being the arrival of additional data. However, this distribution can in addition be updated successively for the same data set, as has also been pointed out to us by Arnoldo Frigessi. We call these

versions of the algorithm Parametric and Non Parametric Sequential Adaptive Independent Chain (PSAIC and NPSAIC). Note that since now $\boldsymbol{\theta}'$, and hence $\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}')$, depends on all previous iterations on the same data set, the Markov property is destroyed. Nevertheless, one would intuitively expect the algorithm to converge. This seems to be confirmed at least for the implementation of the PSAIC algorithm in Section 4. A proof of convergence for a modified version of this algorithm is given in Appendix 2. It is also possible to update the proposal distribution successively for the complete data set only. We have chosen to do this when a sequence of a fixed number of samples is completed. We call these versions of the algorithm Parametric and Non Parametric Adaptive Independent Chain (PAIC and NPAIC).

When running the PSIC, PSAIC and PAIC we use a simple diagnostic test of convergence. For each l we choose threshold values $\varepsilon_{a_l}, \varepsilon_{b_l} > 0$ corresponding to (a_l, b_l) the shape and scale parameter of the gamma marginal proposal distribution of θ_l. Considering PAIC, when the difference between the parameter estimates based on two consecutive sequences becomes less than these thresholds for both scale and shape parameter and for each l, the burn-in is terminated and the proposal distribution is kept fixed at the last updated value for the rest of the iterations. For the PSAIC we follow the same procedure, using the same criterion to determine when to take in a new data point. The latter is done also for PSIC. The test does of course not ensure that stationarity is reached, but at any rate it normally ensures that the proposal distribution has stabilized, and our numerical examples indicate that the test works well in practice.

In Sections 3 and 4 we will compare the various types of sequential Metropolis-Hastings algorithms with standard simulation techniques such as Rejection Sampling, ordinary Metropolis-Hastings and the Sampling Importance Resampling (SIR) algorithm. For the latter algorithm see Smith and Gelfand (1992) where it is called Weighted Bootstrap. To apply rejection sampling in the Marshall-Olkin shock model we must find an upper bound on $L(\boldsymbol{\theta})$ given by (1.4), again see Smith and Gelfand (1992). Remembering that C_A is the set of critical shocks for A, we establish the following upper bound

$$\begin{aligned} L(\boldsymbol{\theta}) &= \lim_{dt \to 0} P(t < T \leq t + dt, D = A)/dt \\ &= \sum_{l \in C_A} \theta_l \exp(-\theta_l t) \lim_{dt \to 0} P(t < T \leq t + dt, D = A \mid t \leq V_l < t + dt) \\ &\leq \sum_{l \in C_A} \theta_l exp(-\theta_l t) \leq |C_A|/(te) \end{aligned}$$

This general upper bound can be improved in specific applications.

3 An application to preventive system maintenance

In this section, following Gåsemyr and Natvig (1999), we consider preventive system maintenance where components are replaced according to a specific strategy. We have to take into account that it is costly to intervene in system operation. Hence, it is desirable to postpone replacement of failed components as long as possible in order to replace several components at a time. On the other hand, it is obviously important to avoid a system failure. As a compromise we assume that components are replaced as soon as system weakening has reached a certain level; i.e. when $\psi(\boldsymbol{X}(t))$ jumps to zero, where ψ is a binary, monotone structure function such that $\psi(\boldsymbol{X}(t)) \leq \phi((\boldsymbol{X}(t))$. At this time a total inspection of the components is carried through and all failed components are replaced, while the others are not affected. We assume this procedure takes zero operational time. Afterwards, the replaced components are assumed to have the same lifetime distributions as the initial ones. It is natural to choose ψ such that when $\psi(\boldsymbol{X}(t))$ jumps to zero, at least one additional component must fail for $\phi(\boldsymbol{X}(t))$ to jump to zero.

Consider the network system of seven components given in Figure 1.

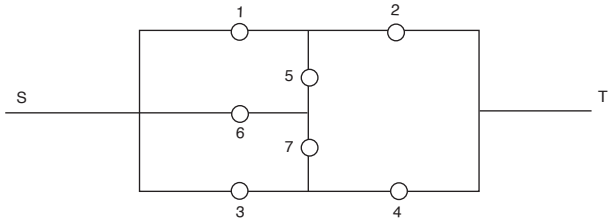

Figure 1: Network system of seven components

The system is working iff there is at least one connection between S(ource) and T(erminal). In Gåsemyr and Natvig (1999) it is shown that a natural choice for $\psi(\boldsymbol{x})$ is:

$$\psi(\boldsymbol{x}) = x_2x_4[x_1x_3 + (1 - x_3)x_1x_6x_7 + (1 - x_1)x_3x_5x_6]$$

We have simulated 20 exponentially distributed lifetimes for each of the components $1, \ldots, 7$ with expectations θ_l^{-1}, $l = 1, \ldots, 7$ measured in hours respectively equal to 1000, 2000, 1800, 1500, 600, 800, 700. In Gåsemyr and Natvig (1999) the following likelihood function, $L(\boldsymbol{\theta})$, is established, based on data from observing the system components according to a specifically described scheme on the interval $[0, 10000]$:

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \theta_1^{11} e^{-9395\theta_1} \theta_2^3 e^{-10000\theta_2} \theta_3^5 e^{-9128\theta_3} \theta_4^5 e^{-10000\theta_4} \\
&\times \theta_5^2 e^{-4497\theta_5} (1-e^{-449\theta_5})(1-e^{-273\theta_5})(1-e^{-447\theta_5})(1-e^{-337\theta_5})(1-e^{-1614\theta_5}) \\
&\times (1-e^{-860\theta_5})(1-e^{-1223\theta_5})(1-e^{-295\theta_5}) \\
&\times \theta_6 e^{-2904\theta_6} (1-e^{-866\theta_6})(1-e^{-346\theta_6})(1-e^{-253\theta_6})(1-e^{-499\theta_6})(1-e^{-449\theta_6})(1-e^{-115\theta_6}) \\
&\times (1-e^{-337\theta_6})(1-e^{-1614\theta_6})(1-e^{-1223\theta_6})(1-e^{-284\theta_6})(1-e^{-810\theta_6})(1-e^{-295\theta_6}) \\
&\times \theta_7 e^{-3682\theta_7} (1-e^{-866\theta_7})(1-e^{-346\theta_7})(1-e^{-499\theta_7})(1-e^{-449\theta_7})(1-e^{-337\theta_7}) \\
&\times (1-e^{-1614\theta_7})(1-e^{-1223\theta_7})(1-e^{-284\theta_7})(1-e^{-299\theta_7})(1-e^{-396\theta_7})
\end{aligned} \tag{3.1}
$$

Assume that we choose the failure rates $\theta_1, \ldots, \theta_7$ to be independent à priori each having an ordinary gamma distribution $g(\theta; 1, 1000)$ with expectation and standard deviation equal to 0.001. From (3.1) $\theta_1, \ldots, \theta_7$ are independent à posteriori as well. The posterior distribution of θ_l is $g(\theta_l; a_l, b_l)$, $l = 1, 2, 3, 4$ with (a_l, b_l) respectively equal to $(12, 10395)$, $(4, 11000)$, $(6, 10128)$ and $(6, 11000)$. For $l = 5, 6, 7$ the posterior distribution is $g(\theta_l; a_l, b_l, \boldsymbol{t}_l)$ with $(a_l, b_l, \boldsymbol{t}_l)$ respectively equal to

$$
\begin{aligned}
&(3, 5497, 449, 273, 447, 337, 1614, 860, 1223, 295)\ , \\
&(2, 3904, 866, 346, 253, 499, 449, 115, 337, 1614, 1223, 284, 810, 295) \\
&(2, 4682, 866, 346, 499, 449, 337, 1614, 1223, 284, 299, 396)
\end{aligned}
$$

To go through all 2^m addends in (1.2) the following algorithm is used. Let

$$
A(1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad A(i) = \begin{pmatrix} \mathbf{0} & A(i-1) \\ \mathbf{1} & A(i-1) \end{pmatrix}\ , \qquad i = 2, \ldots, m
$$

Here $\mathbf{0}$ and $\mathbf{1}$ are column vectors with dimension 2^{i-1}. $A(i)$ is a $2^i \times i$ matrix. Hence $A(m)$ gives all 2^m desired combinations as row vectors.

In the same figures we have plotted (full lines) the corresponding simulated posterior distributions. These are based on Rejection Sampling (RS), Parametric Independent Chain (PIC), an ordinary random walk Metropolis-Hastings algorithm which we call Parametric Dependent Chain (PDC) using a truncated normal proposal density, Sampling Importance Resampling (SIR) and finally the Parametric and Non Parametric Adaptive Independent Chain (PAIC and NPAIC). Since we consider the data given in this application as one data set, sequential versions of the two latter algorithms have not been considered. Inspecting the figures, especially Figure 3 covering θ_6, the RS algorithm works poorly. It is way out for $1.5 \cdot 10^{-3} < \theta_6 < 2.5 \cdot 10^{-3}$.

All the rest do rather well. It should be noted that θ_6 has the largest number of left censored lifetimes entering into the posterior ($m = 12$). Hence this posterior is the most difficult to calculate.

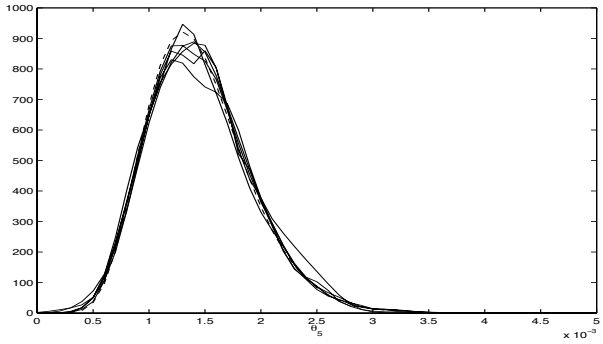

Figure 2: The densities for θ_5 for the analytical posterior (dashed) and simulated posteriors (full lines) for all simulation algorithms.

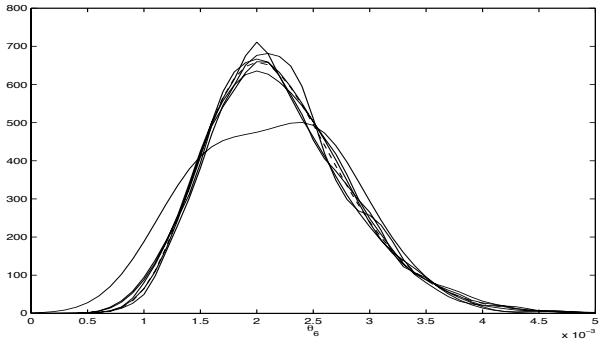

Figure 3: The densities for θ_6 for the analytical posterior (dashed) and simulated posteriors (full lines) for all simulation algorithms.

In Table 1 we have given the execution times (in seconds) for all algorithms to make a proper comparison.

For the RS algorithm that works poorly execution time is very long. Obviously, a much longer execution time would have been needed to make it work. The PAIC algorithm is the fastest, closely followed by PIC. SIR is a clear number three. Then follows PDC, whereas NPAIC is definitely the slowest among the best five.

The execution times listed in Table 1 are the results of a single run for each algorithm, and hence do not represent a thorough numerical experiment. Nevertheless, the results for the three different parameters show a fairly consistent pattern, and provide a reasonable indication of the speed

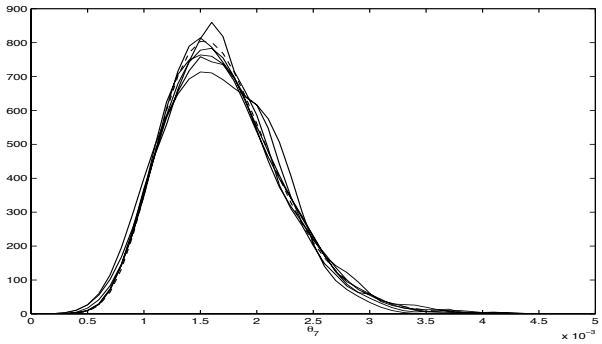

Figure 4: The densities for θ_7 for the analytical posterior (dashed) and simulated posteriors (full lines) for all simulation algorithms.

Table 1. Execution times in seconds for all algorithms for θ_l, $l = 5, 6, 7$.

	RS	PIC	PDC	SIR	PAIC	NPAIC
θ_5	1077	20	298	37	15	899
θ_6	1214	24	302	185	16	1351
θ_7	1147	22	301	78	19	1110

of the algorithms. The tuning of the algorithms was done after some trial runs and was done with the intention to obtain similar quality for the density approximations for all algorithms. The quality was assessed by visual inspection; no formal distance measure for densities was used. A rule of thumb emerging from our trials was that a total number of somewhere between 2000 and 3000 different values in the final sample, which means after burn-in for the MCMC algorithms, were needed in order for the density estimate to produce acceptable results. Hence the total number of iterations in the MCMC algorithms after burn-in had to be roughly inversely proportional to the acceptance rate.

In Table 2 we have listed some key figures from the computations for all simulation algorithms. Due to the very long execution times, we did not wait for the RS to produce reasonable values for the number of accepted proposals.

For RS, PIC, SIR and PAIC the prior gamma, $g(\theta; 1, 1000)$, is used as proposal distribution. For PDC and the first sequence of NPAIC a normal density, restricted to $[0, 0.01]$, with expectation equal to the previous sample and variance equal to 10^{-4} is used as proposal distribution. For PAIC a sequence length of 100 is applied, and we choose $\varepsilon_{a_l} = 0.2$, $\varepsilon_{b_l} = 5$, $l = 5, 6, 7$.

Table 2. Key figures from the computations for all algorithms for θ_l, $l = 5, 6, 7$.

		Length of burn-in	Total number of iterations after burn-in	Total number of accepted proposals after burn-in/resamples*
RS	θ_5	–	1200000	194
	θ_6	–	1200000	13
	θ_7	–	1200000	29
PIC	θ_5	3000	17000	5661
	θ_6	3000	17000	3708
	θ_7	3000	17000	4873
PDC	θ_5	2000	18000	2847
	θ_6	2000	18000	3661
	θ_7	2000	18000	3093
SIR	θ_5	–	40000	2000*
	θ_6	–	180000	2000*
	θ_7	–	80000	2000*
PAIC	θ_5	2300	3000	2944
	θ_6	2200	3000	2888
	θ_7	3200	3000	2946
NPAIC	θ_5	2000	3000	2687
	θ_6	3000	3000	2816
	θ_7	2500	3000	2766

4 An application to a simple shock model

In this section, following Gåsemyr and Natvig (1998), we consider a simple parallel system of two components subjected to a common shock. This is a special case of the general shock model presented in Section 1 with $E = \{1, 2\}$ and $S = \{1, 2, 3\}$. The only fatal set is $A = \{1, 2\}$. Suppose the prior distribution for $\boldsymbol{\theta}$ is given by (1.7), where (a_l, b_l), $l = 1, 2, 3$ are respectively given by (4,1), (6,3) and (6,2). The likelihood function (1.4) now obviously reduces to:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \theta_1(1 - \exp(-\theta_2 t)) \exp(-(\theta_1 + \theta_3)t) \\ &+\theta_2(1 - \exp(-\theta_1 t)) \exp(-(\theta_2 + \theta_3)t) + \theta_3(1 - \exp(-\theta_2 t)) \exp(-(\theta_1 + \theta_3)t) \\ &+\theta_3(1 - \exp(-\theta_1 t)) \exp(-(\theta_2 + \theta_3)t) + \theta_3 \exp(-(\theta_1 + \theta_2 + \theta_3)t) \end{aligned} \tag{4.1}$$

In Gåsemyr and Natvig (1998) an explicit expression for the posterior distribution $\pi(\boldsymbol{\theta}|t, A)$ is given whereas $\pi(\boldsymbol{\theta}|t_1, A, t_2, A)$ is given in Sørensen

(1999). For autopsy data from more than two independent systems no explicit analytical posterior has been calculated. The analytical prior and posterior distributions for θ_l, $l = 1, 2, 3$ are plotted respectively in Figures 5, 6, 7 for $\boldsymbol{t} = (1.5, 1.5)$ and in Figures 8, 9, 10 for $\boldsymbol{t} = (3, 3)$.

In the same figures we have as in Section 3 plotted the corresponding simulated posterior distributions. These are based on RS, except for $\boldsymbol{t} = (3, 3)$, two ordinary Metropolis-Hastings algorithms, PIC and PDC, SIR, and finally PAIC and three sequential Metropolis-Hastings algorithms PSAIC, PSIC and NPSIC. For $\boldsymbol{t} = (1.5, 1.5, 0.4, 0.8, 0.5)$ the simulated posterior distributions for θ_l, $l = 1, 2, 3$, based on the same algorithms, are plotted in Figures 11, 12, 13.

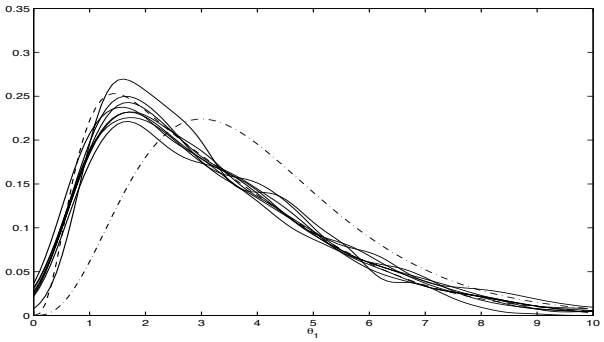

Figure 5: The densities for θ_1 for the prior $(-\cdot)$, analytical posterior $(- -)$ and simulated posteriors (full lines) for all simulation algorithms, with $\boldsymbol{t} = (1.5, 1.5)$.

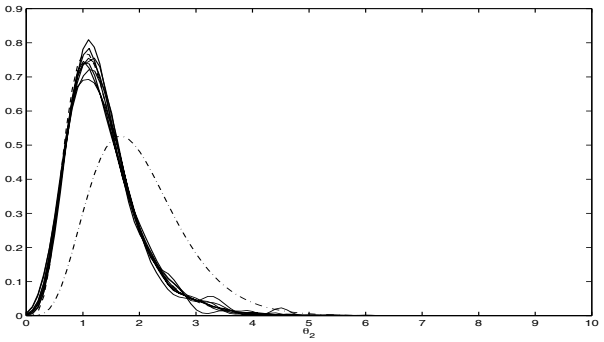

Figure 6: The densities for θ_2 for the prior $(-\cdot)$, analytical posterior $(- -)$ and simulated posteriors (full lines) for all simulation algorithms, with $\boldsymbol{t} = (1.5, 1.5)$.

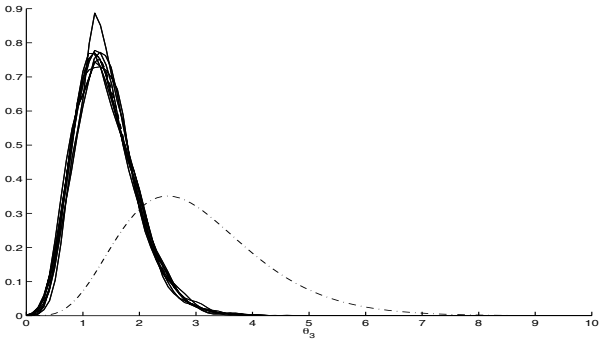

Figure 7: The densities for θ_3 for the prior $(-\cdot)$, analytical posterior $(- -)$ and simulated posteriors (full lines) for all simulation algorithms, with $\boldsymbol{t} = (1.5, 1.5)$.

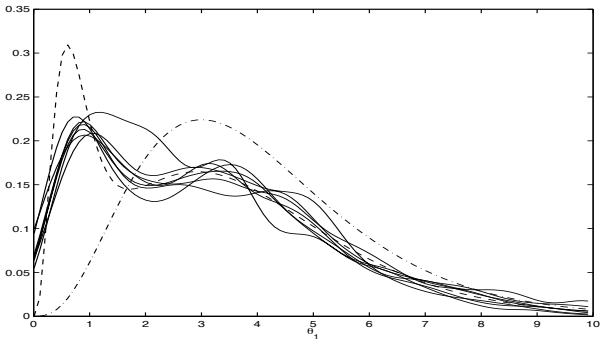

Figure 8: The densities for θ_1 for the prior $(-\cdot)$, analytical posterior $(- -)$ and simulated posteriors (full lines) for all simulation algorithms except RS, with $\boldsymbol{t} = (3, 3)$.

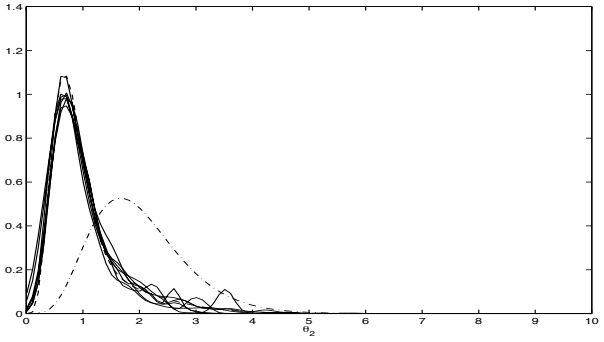

Figure 9: The densities for θ_2 for the prior $(-\cdot)$, analytical posterior $(- -)$ and simulated posteriors (full lines) for all simulation algorithms except RS, with $\boldsymbol{t} = (3, 3)$.

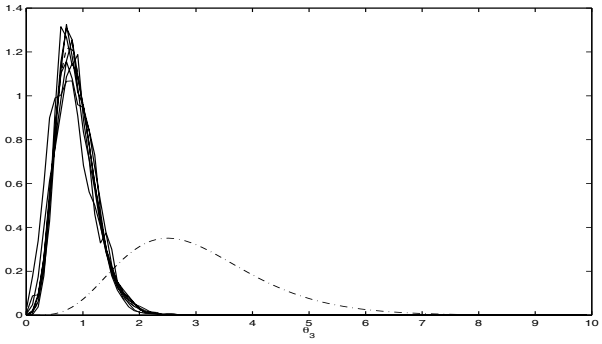

Figure 10: The densities for θ_3 for the prior $(-\cdot)$, analytical posterior $(- -)$ and simulated posteriors (full lines) for all simulation algorithms except RS, with $\boldsymbol{t} = (3, 3)$.

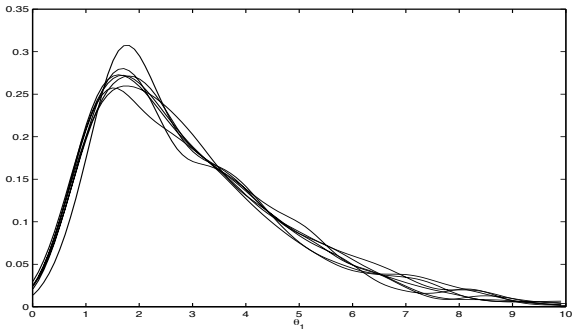

Figure 11: The densities for θ_1 for the simulated posteriors (full line) for all simulation algorithms, with $\boldsymbol{t} = (1.5, 1.5, 0.4, 0.8, 0.5)$.

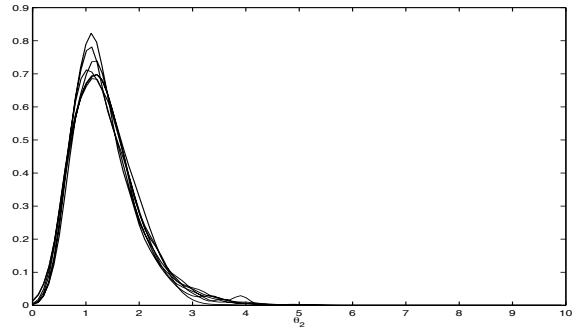

Figure 12: The densities for θ_2 for the simulated posteriors (full line) for all simulation algorithms, with $\boldsymbol{t} = (1.5, 1.5, 0.4, 0.8, 0.5)$.

Inspecting Figures 5, 6, 7, comparing with the analytical posterior, we see that all algorithms do rather well for θ_l, $l = 1, 2, 3$ when $\boldsymbol{t} = (1.5, 1.5)$. However, when $\boldsymbol{t} = (3, 3)$, where we see from Figures 8, 9, 10 that there are large discrepancies between the prior assessments and the data, it turns out that RS does not work at all. We see from Figures 9, 10 that for θ_2 and θ_3 the remaining algorithms do rather well. For θ_1 from Figure 8 and inspecting the individual posteriors, plotted in figures not presented here, NPSIC works poorly, whereas the rest of the algorithms do not work that badly.

Finally, we inspect Figures 11, 12, 13, covering $\boldsymbol{t} = (1.5, 1.5, 0.4, 0.8, 0.5)$, where no analytical posterior has been calculated. This would have involved $5^5 = 3125$ addends in the joint posterior distribution for $\boldsymbol{\theta}$ due to 5 addends in (4.1) and autopsy data from 5 independent systems. We see from these figures that the different plots are reasonably close for θ_l, $l = 1, 2, 3$ indicating that all algorithms do rather well in this case. This is not surprising remembering that this was true for $\boldsymbol{t} = (1.5, 1.5)$ and noting that the rest of the data are in accordance with prior assessments.

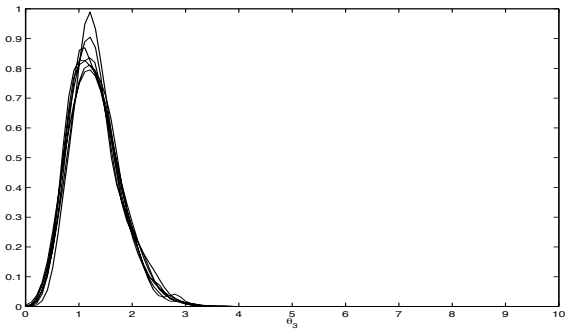

Figure 13: The densities for θ_3 for the simulated posteriors (full line) for all simulation algorithms, with $\boldsymbol{t} = (1.5, 1.5, 0.4, 0.8, 0.5)$.

The tuning of the algorithms was done along the same lines as in Section 3. In Table 3 we have given the execution times (in seconds), and also the total number of iterations including burn-in, to arrive at the simulated posteriors for all θ_l, $l = 1, 2, 3$, for all algorithms to make a proper comparison.

Table 3: Execution times in seconds, and total number of iterations including burn-in, to arrive at the simulated posteriors for all θ_l, $l = 1, 2, 3$, for all algorithms.

	$\boldsymbol{t}$	Execution times (in seconds)	Total number of iterations including burn-in
RS	(1.5,1.5)	1361	1017234
	(3,3)	–	–
	(1.5,1.5,0.4,0.8,0.5)	3125	3872491
PIC	(1.5,1.5)	101	50000
	(3,3)	514	250000
	(1.5,1.5,0.4,0.8,0.5)	153	50000
PDC	(1.5,1.5)	464	10000
	(3,3)	461	10000
	(1.5,1.5,0.4,0.8,0.5)	237	5000
SIR	(1.5,1.5)	46	32000
	(3,3)	279	200000
	(1.5,1.5,0.4,0.8,0.5)	46	32000
PAIC	(1.5,1.5)	53	5200
	(3,3)	104	8800
	(1.5,1.5,0.4,0.8,0.5)	103	6400
PSAIC	(1.5,1.5)	49	5000
	(3,3)	93	8200
	(1.5,1.5,0.4,0.8,0.5)	86	8600
PSIC	(1.5,1.5)	89	8700
	(3,3)	114	10200
	(1.5,1.5,0.4,0.8,0.5)	141	14000
NPSIC	(1.5,1.5)	1892	6000
	(3,3)	1900	6000
	(1.5,1.5,0.4,0.8,0.5)	1674	4900

For RS that does not work at all for $\boldsymbol{t} = (3, 3)$ and for NPSIC that works poorly for θ_1 for this data vector, execution times are extremely long. The PSAIC algorithm is the fastest followed by PAIC. In the next group follow PSIC and SIR, whereas the two ordinary Metropolis-Hastings algorithms PIC and PDC are definitely the slowest among the best six.

For RS, PIC, SIR, PAIC, PSAIC, PSIC the product of prior gammas is used as the initial proposal distribution. For PDC and for the processing of the first data point of NPSIC a normal density, restricted to $[0, \infty]$, with expectation equal to the previous sample and variance equal to 1 is used as proposal distribution. For PAIC, PSAIC, PSIC a sequence length of 200 is applied, and we choose $\varepsilon_{a_l} = \varepsilon_{b_l} = 0.2$, $l = 1, 2, 3$ as threshold values

to arrive at simulated posteriors for all θ_l, $l = 1, 2, 3$. Considering $\boldsymbol{t} = (1.5, 1.5, 0.4, 0.8, 0.5)$ for PAIC the diagnostic test concluded the iterations after 17 sequences. The corresponding numbers of sequences were 3, 5, 7, 9 and 4, altogether 28, for PSAIC and 8, 11, 10, 16 and 10, altogether 55, for PSIC. Hence the lengths of burn-in for this data vector were respectively 3400, 5600 and 11000 for these algorithms.

5 Concluding remarks

In this paper focus has been on computational aspects of some models in reliability, presented in Gåsemyr and Natvig (1998, 1999), involving the generalized gamma distribution. We have tried several simulation algorithms. Among these are two new types of sequential Metropolis-Hastings algorithms introduced here. We have used artificial data on examples taken from Gåsemyr and Natvig (1998, 1999). By the variation in data, light is shed on the flexibility of the algorithms. In most of the trials, exact analytical solutions were available as a basis for comparison. A quality criterion has been the ability of the algorithms to reproduce marginal posterior densities close to the true curves in reasonable computation time.

Among the algorithms based on sampling from a fixed distribution, SIR seems to be the fastest and most flexible. RS has a very limited applicability, whereas PIC works well if the prior distribution is in reasonable accordance with the data. However, even SIR has trouble coping with large discrepancies between the prior assessments and the data. The new parametric, sequential Metropolis-Hastings algorithms, PAIC, PSAIC, PSIC generalizing PIC, are superior in coping with such discrepancies, but are also among the faster algorithms in the less problematic cases. Anyway, they work substantially better than the PDC.

We would guess that the parametric, sequential algorithms are even more favourable in more complex cases. More complicated proposal distributions, such as convex combinations of products of gamma or generalized gamma distributions may then be useful.

The idea behind the parametric, sequential algorithms is quite general and is potentially useful in may other situations, where other parametric classes of proposal distributions may be more natural, e.g. normal distributions or combinations of uniform distributions. We hope that this paper might stimulate research in this direction.

The non parametric, sequential algorithm, NPSIC tried out in this pa-

per, does not look too promising so far, due to the very long computation times. This type of algorithm might be useful, however, if it is difficult to find a parametric class that works.

Acknowledgement

This research has partly been supported by the Norwegian Research Council under strategic industrial project no. 121144/420 and Norwegian Computing Center project no. 830100. We are very thankful for this support. We are especially thankful to Arnoldo Frigessi at the Norwegian Computing Center for helpful comments and ideas.

We will also express our gratitude to our colleagues Arne Bang Huseby and Geir Storvik for respectively giving a comment to the present form of the generalized gamma distribution and for helpful discussions on simulation procedures based on the Metropolis-Hastings algorithm.

References

[1] Besag, J., Green, P., Higdon, D. and Mengersen, K., "Bayesian computation and stochastic systems," Statist. Science 10, 3–66, 1995.

[2] Bølviken, E., Christophersen, N. and Storvik, G., Linear dynamical models, Kalman filtering and statistics, Lecture notes. Departments of Mathematics and Informatics, University of Oslo, 1998.

[3] Chib, S. and Greenberg, E., "Understanding the Metropolis-Hastings algorithm," Amer. Statist. 49, 327–335, 1995.

[4] Gåsemyr, J. and Natvig, B., "Some aspects of reliability analysis in shock models," Scand. J. Statist. 22, 385–393, 1995.

[5] Gåsemyr, J. and Natvig, B. "The posterior distribution of the parameters of component lifetimes based on autopsy data in a shock model," Scand. J. Statist. 25, 271–292, 1998.

[6] Gåsemyr, J. and Natvig, B., "Bayesian inference based on partial monitoring of components with applications to preventive system maintenance," Submitted Lifetime Data Analysis, 1999.

[7] Lindvall, T., Lectures on the coupling method. Wiley, New York, 1992.

[8] Marshall, A.W. and Olkin, I., "A multivariate exponential distribution," J. Amer. Statist. Ass. 62, 30–44, 1967.

[9] Meilijson, I., "Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine," J. Appl. Prob. 18, 829–838, 1981.

[10] Smith, A.F.M. and Gelfand, A.E., "Bayesian statistics without tears: a sampling-resampling perspective," Amer. Statist. 46, 84–88, 1992.

[11] Smith, A.F.M. and Roberts, G.O., "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods," J. R. Statist. Soc. B. 55, 3–23, 53–102, 1993.

[12] Sørensen, E., Simulating the posterior distribution in reliability models involving the generalized gamma distribution. Cand. scient. thesis. Department of Mathematics, University of Oslo, 1999.

[13] Tierney, L., "Markov chains for exploring posterior distributions," Ann. Statist. 22, 1701–1762, 1994.

Appendix 1

A comparison of the representations (1.5) and (1.9) with respect to computational efficiency

Starting out with (1.5), let $\boldsymbol{t} = (t_1, t_2, \cdots, t_r)$ be a vector with positive entries such that $\boldsymbol{t}_{j,l}$ is a subvector of $\boldsymbol{t}$ for all $j = 1, \cdots, J$, $l = 1, \cdots, n+p$. For $j = 1, \cdots, J$, $l = 1, \cdots, n+p$ define $\boldsymbol{d}_{j,l} = (d_{j,l,1}, d_{j,l,2}, \cdots, d_{j,l,r})$ by putting $d_{j,l,i} = 1$ if t_i occurs as an entry in the subvector $\boldsymbol{t}_{j,l}$ and 0 otherwise. Note that the normalizing constant corresponding to the lth factor of the jth summand of (1.5) can be calculated as in (1.2)

$$\gamma(a_{j,l}, b_{j,l}, \boldsymbol{t}_{j,l})^{-1} = (\Gamma(a_{j,l})) \sum_{\{\boldsymbol{d} \in (0,1)^r | \boldsymbol{d} \leq \boldsymbol{d}_{j,l}\}} (b_{j,l} + \boldsymbol{d} \cdot \boldsymbol{t})^{-a_{j,l}} \tag{A1.1}$$

Introduce $\boldsymbol{d}_j = (d_{j,1,1}, \ldots, d_{j,1,r}, \ldots, d_{j,n+p,1}, \ldots, d_{j,n+p,r})$, i.e. the vector made up by the subvectors $\boldsymbol{d}_{j,l}$, $l = 1, \cdots, n+p$. The distribution in (1.5) can then be written in the form (1.9) as

$$\begin{aligned} \pi(\boldsymbol{\theta}) &\propto \sum_{j=1}^{J} \prod_{l=1}^{n+p} \prod_{i=1}^{r} (1 - \exp(-\theta_l t_i))^{d_{j,l,i}} h(\theta_l; a_{j,l}, b_{j,l}) \\ &= \sum_{j=1}^{J} \sum_{\{\boldsymbol{c} \in \{0,1\}^{r(n+p)} | \boldsymbol{c} \leq \boldsymbol{d}_j\}} (-1)^{|\boldsymbol{c}|} \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l} + \sum_{i=1}^{r} c_{(l-1)r+i} t_i) \end{aligned} \tag{A1.2}$$

To calculate the distribution from (A1.2), one must calculate a weight for each of the $\sum_{j=1}^{J} 2^{|\boldsymbol{d}_j|}$ summands, each of which involves a product of $n+p$ factors of the form $\Gamma(a)(b+\boldsymbol{d}\cdot\boldsymbol{t})^{-a}$. In contrast, (1.5) has only J summands, and the weight corresponding to the jth summand involves $\sum_{l=1}^{n+p} 2^{|\boldsymbol{d}_{j,l}|}$ terms of the form $\Gamma(a)(b + \boldsymbol{d} \cdot \boldsymbol{t})^{-a}$ (see (A1.1)). Note that

$$\sum_{l=1}^{n+p} |\boldsymbol{d}_{j,l}| = \sum_{l=1}^{n+p} \sum_{i=1}^{r} d_{j,l,i} = |\boldsymbol{d}_j|$$

Hence,

$$\sum_{l=1}^{n+p} 2^{|\boldsymbol{d}_{j,l}|} \leq 2^{|\boldsymbol{d}_j|}$$

It is difficult to give a precise comparison of the computational complexity involved in the two different forms, since this will depend on concrete implementations of computation algorithms. It may be possible to make efficient use of the fact that different weights contain many identical factors. Nevertheless, the above discussion indicates strongly that (1.5) is considerably more efficient computationally than (1.9) in the present model, thus providing a good case for the usefulness of the generalized gamma distribution.

Appendix 2

A proof of convergence for a modified version of the PSAIC algorithm

The following notation is used in the proof. r denotes the total number of data points; π is the posterior distribution based on all r data points, i.e. the final target distribution; the length of sequences between updating of the proposal distribution is denoted by S; (a_i, b_i), $i = 1, 2, 3$ are the parameters of the original prior gamma distributions; (a_i^u, b_i^u) are the last updates of (a_i, b_i) based on estimates of $E(\theta_i)$ and $\mathrm{Var}(\theta_i)$, $i = 1, 2, 3$ using the last sequence of S iterations. Finally, g^u denotes the corresponding product of gammas proposal distribution.

The PSAIC algorithm is modified as follows. In the first place, the total number of sequences before adding the last rth data point is bounded, either by using a fixed number of sequences for each data point or by a combination with the diagnostic test described in Section 2. Secondly, (a_i^u, b_i^u) is restricted to the set $A_i = [0, a_i] \times [\delta, b_i - \delta]$ for some $0 < \delta < b_i/2$, $i = 1, 2, 3$.

Now, let M be the number of iterations before the last rth data point is added. M may be random, but is bounded. Let $\boldsymbol{\eta}_M$ be a random variable with distribution π and let $\{\boldsymbol{\eta}_k\}_{k \geq M}$ be a stochastic process with $\boldsymbol{\eta}_M$ as initial value. This process moves according to a Metropolis-Hastings algorithm with the same proposal distribution as the $\{\boldsymbol{\theta}_k\}$ chain. The two chains are linked by using the same proposed values $\boldsymbol{\theta}'$ the acceptance of which occurs in each chain if a common uniform U_k is less than the corresponding acceptance probability. Clearly, the $\{\boldsymbol{\eta}_k\}_{k \geq M}$ chain starting in the stationary distribution π, remains π-distributed throughout the first sequence of S iterations. In particular $\boldsymbol{\eta}_{M+S}$ is π-distributed. Now $\boldsymbol{\eta}_{M+S}$ can be regarded as the initial value for a new chain with an updated proposal distribution. Repeating the argument above we conclude that $\boldsymbol{\eta}_k$ is π-distributed for all $k \geq M$.

Let now $K = \min\{k : \boldsymbol{\eta}_k = \boldsymbol{\theta}_k\}$ be the coupling time of the two chains. Note that starting from $(\boldsymbol{\eta}_k, \boldsymbol{\theta}_k)$ and proposing $\boldsymbol{\theta}'$, due to the common uniform U_k, coupling occurs with probability

$$\begin{aligned}\min\{1, \pi(\boldsymbol{\theta}')g^u(\boldsymbol{\theta}_k)/(\pi(\boldsymbol{\theta}_k)g^u(\boldsymbol{\theta}')), \pi(\boldsymbol{\theta}')g^u(\boldsymbol{\eta}_k)/(\pi(\boldsymbol{\eta}_k)g^u(\boldsymbol{\theta}'))\} \\ \geq \pi(\boldsymbol{\theta}')/(g^u(\boldsymbol{\theta}')w^u) \ ,\end{aligned}$$

where w^u is the supremum of $\pi(\boldsymbol{\theta})/g^u(\boldsymbol{\theta})$ with $\boldsymbol{\theta}$ ranging over $[0, \infty)^3$. In-

tegrating with respect to $g^u(\boldsymbol{\theta}')$, we find that the coupling probability is bounded from below by $1/w^u$. Now let w be a common upper bound for w^u as g^u ranges over possible proposals having parameters (a_i^u, b_i^u) restricted to the set A_i, $i = 1, 2, 3$. Since from Section 4 the likelihood contribution to π, $\prod_{i=1}^{r} L(\boldsymbol{\theta}) \mid t_i, A_{j_i})$, is bounded, it then follows that the upper bound does in fact exist. Furthermore, it follows that $K - M$ is stochastically dominated by a geometrically distributed variable with parameter $1/w$. From the coupling inequality given in Lindvall (1992) it finally follows that the distribution of $\{\boldsymbol{\theta}_k\}$ converges to π in the total variation norm.