

ON A GENERALIZATION OF THE GAMMA DISTRIBUTION

Jørund Gåsemyr and Bent Natvig

University of Oslo

Abstract

In this paper we define a generalized gamma distribution, which is shown to enter naturally in Bayesian analysis of exponential survival models with left censoring. Some simple examples are given in Section 1. In Section 2 we discuss the use of this distribution in the analysis of autopsy data in a shock model, an application which is treated in depth in Gåsemyr & Natvig (1995b). In the present paper we also focus on a comparison of the generalized gamma distribution with the ordinary one with respect to computational efficiency in this model. Furthermore, we suggest in Section 3 simulation procedures mostly based on the Metropolis-Hastings algorithm that may be used for parameter estimation and prediction in situations where exact methods are intractable.

Key words: exponential survival models, left censoring, autopsy data, shock model, simulation, Metropolis - Hastings algorithm, adaptive sampler.

1 The generalized gamma distribution

It is a well known fact that the class of gamma distributions is a natural conjugate class of prior distributions for the exponential distribution. This is the case also for exponential models which may lead to right censored observations. However, in some models left censoring arises naturally. In this note we introduce a generalized gamma distribution, which turns out to be useful in order to handle such models. We also give some simple examples on the use of the distribution in this section, and elaborate on the more complicated application to the exponential autopsy shock model, considered in Gåsemyr & Natvig (1995b), in Section 2. In Section 3 we suggest simulation procedures mostly based on the Metropolis - Hastings algorithm that may be used for parameter estimation and prediction in situations where exact methods are intractable.

For positive real numbers a, b, t_1, \dots, t_m , $m \geq 0$, define the functions

$$h(\theta; a, b, \underline{t}) = \theta^{a-1} e^{-b\theta} \prod_{i=1}^m (1 - e^{-\theta t_i}), \quad \theta \geq 0, \quad (1.1)$$

where $\underline{t} = (t_1, \dots, t_m)$. Corresponding to $m = 0$, we have in particular $h(\theta; a, b) =$

$\theta^{a-1}e^{-b\theta}$. Define the normalizing constant $\gamma(a, b, \underline{t})$ by

$$\begin{aligned}
(\gamma(a, b, \underline{t}))^{-1} &= \int_0^\infty h(\theta; a, b, \underline{t}) d\theta \\
&= \sum_{B \subset \{1, \dots, m\}} (-1)^{|B|} \int_0^\infty h(\theta; a, b + \sum_{i=1}^m I_B(i) t_i) d\theta \\
&= \sum_{\underline{d} \in \{0, 1\}^m} (-1)^{|\underline{d}|} \int_0^\infty h(\theta; a, b + \underline{d} \cdot \underline{t}) d\theta \\
&= \Gamma(a) \sum_{\underline{d} \in \{0, 1\}^m} (-1)^{|\underline{d}|} (b + \underline{d} \cdot \underline{t})^{-a},
\end{aligned} \tag{1.2}$$

where $|B|$ denotes the number of elements in B , $d_i = I_B(i)$ and $|\underline{d}|$ denotes $d_1 + \dots + d_m$. The generalized gamma distribution with parameters a, b and \underline{t} is then defined as the probability distribution on $[0, \infty)$ with density function given by

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

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

As seen by the subsequent examples, the generalized gamma distribution arises naturally as the posterior distribution for the failure rate θ in certain exponential models. One will then often be interested in estimating θ . The standard Bayes estimator, minimizing the expected loss with respect to a quadratic loss function, is

$$E(\theta) = \gamma(a, b, \underline{t}) / \gamma(a + 1, b, \underline{t}) = a \frac{\sum_{\underline{d} \in \{0, 1\}^m} (-1)^{|\underline{d}|} (b + \underline{d} \cdot \underline{t})^{-(a+1)}}{\sum_{\underline{d} \in \{0, 1\}^m} (-1)^{|\underline{d}|} (b + \underline{d} \cdot \underline{t})^{-a}} \tag{1.4}$$

Another quantity of interest is the predictive survival probability beyond some time t for a variable T , which, given θ , is exponentially distributed with failure rate θ . This is given by

$$\begin{aligned}
P(T > t) &= \int_0^\infty P(T > t | \theta) g(\theta; a, b, \underline{t}) d\theta = \gamma(a, b, \underline{t}) / \gamma(a, b + t, \underline{t}) \\
&= \frac{\sum_{\underline{d} \in \{0, 1\}^m} (-1)^{|\underline{d}|} (b + t + \underline{d} \cdot \underline{t})^{-a}}{\sum_{\underline{d} \in \{0, 1\}^m} (-1)^{|\underline{d}|} (b + \underline{d} \cdot \underline{t})^{-a}}
\end{aligned} \tag{1.5}$$

Example 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 i th component is inspected

at a deterministic inspection time t_i , and it is observed whether the component has failed before t_i . The inspection interferes 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. A priori θ is assumed to be gamma distributed with parameters a, b . Let T_i be the lifetime of the i th component, interpreted as the potential lifetime that would have resulted if the experiment had not been interrupted. Define $D_i = I(T_i \geq t_i)$. The likelihood function for the data $D_i = d_i, i = 1, \dots, r$, is then

$$L(\theta|D_1 = d_1, \dots, D_r = d_r) = \prod_{i=1}^r P(D_i = d_i|\theta) = \prod_{i=1}^r (e^{-\theta t_i})^{d_i} (1 - e^{-\theta t_i})^{1-d_i}$$

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

$$\pi(\theta|D_1 = d_1, \dots, D_r = d_r) = g(\theta; a, b + \underline{d} \cdot \underline{t}, t_{i_1}, \dots, t_{i_m})$$

Example 2. Suppose r identical components are put on test under identical conditions at time 0. The components can be continuously monitored during the test, but monitoring is costly and is therefore restricted to the interval (s, t) . For simplicity, both s and t are assumed deterministic. As in the previous example, we assume a constant failure rate θ with prior distribution $g(\theta; a, b)$. Let T_i be as in Example 1. We observe the random variables $D_i = I(T_i > s)$, $E_i = I(T_i \geq t)$ and $X_i = (1 - D_i)s + (D_i - E_i)T_i + E_it$. Note that D_i and E_i can be expressed in terms of X_i as $D_i = I(X_i > s)$, $E_i = I(X_i \geq t)$, so the likelihood can be expressed as a Radon- Nikodym derivative of $\underline{X} = (X_1, \dots, X_r)$ with respect to the measure (Lebesgue measure $+\delta_t + \delta_s$) r , where δ_t is the Dirac measure at t . For notational convenience, however, we retain the quantities D_i, E_i in the expression for the likelihood, given by

$$\begin{aligned} L(\theta|D_1 = d_1, E_1 = e_1, X_1 = x_1, \dots, D_r = d_r, E_r = e_r, X_r = x_r) \\ = \theta^{\sum_{i=1}^r (d_i - e_i)} (1 - e^{-\theta s})^{\sum_{i=1}^r (1 - d_i)} e^{-\theta(\sum_{i=1}^r d_i x_i)} \end{aligned}$$

By Bayes theorem we obtain the posterior distribution

$$\begin{aligned} \pi(\theta|D_1 = d_1, E_1 = e_1, X_1 = x_1, \dots, D_r = d_r, E_r = e_r, X_r = x_r) \\ = g(\theta; a + \sum_{i=1}^r (d_i - e_i), b + \sum_{i=1}^r d_i x_i, s, \dots, s), \end{aligned}$$

with $\sum_{i=1}^r (1 - d_i)$ copies of s occuring in the last part.

The density function $g(\theta; a, b, \underline{t})$ can also be expressed in terms of ordinary gamma distributions as

$$g(\theta; a, b, \underline{t}) \propto \sum_{\underline{d} \in \{0,1\}^m} (-1)^{|\underline{d}|} (\Gamma(a)/(b + \underline{d} \cdot \underline{t})^a) g(\theta; a, b + \underline{d} \cdot \underline{t}) \quad (1.6)$$

The number of summands in (1.6) equals the number of summands in the normalizing constant for $g(\theta; a, b, \underline{t})$, given by (1.2), indicating that the computational complexity of the two forms are approximately equal. In the more complicated examples of the next section, however, it will be seen that the generalized version of (1.6) may become considerably more complex than the corresponding expression based on generalizing (1.3). In any case the form (1.3) has considerable conceptual advantages. It expresses the density in terms of a single function, all of whose parameters are interpretable as reflecting updating of a prior distribution with experimental data. For instance, with the non-informative prior $\pi(\theta) = \theta^{-1}$, a, b and \underline{t} 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 (1.6), interpretation of the parameters is much more difficult, especially since the sum contains both positive and negative terms. We therefore think that the generalized gamma distribution will be a useful object in Bayesian analysis of exponential survival models.

2 Application to the autopsy model

Rather than a single component, we now consider a binary, monotone system (E, ϕ) of n binary components, $E = \{1, 2, \dots, n\}$. Let T_i be the lifetime of the i th component, and $X_i(t) = I(T_i > t)$ the state of the i th component, functioning or failed, at time t . ϕ is the structure function, a nondecreasing function on $\{0, 1\}^n$, whose value is either 0 (failed) or 1 (functioning). The state of the system at time t is $\phi(X_1(t), \dots, X_n(t))$. The lifetime of the system is $T = \inf\{t | \phi(X_1(t), \dots, X_n(t)) = 0\}$. 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).

We assume, as in Gåsemyr & Natvig (1995b), that the system can be described by a shock model, see Boyles & Samaniego (1984). There exists a set of shocks $S = \{1, 2, \dots, n + p\}$. Here, $1, 2, \dots, n$ represent individual shocks destroying the corresponding component of E , whereas $n + 1, \dots, n + p$ represent common shocks; the l th shock destroying the components in $D_l \subset E$. Similarly, the set of components destroyed by a set B of shocks is denoted by D_B ; i.e. $D_B = \bigcup_{l \in B} D_l$. Let V_l be the time until the l th shock occurs. Thus, $T_i = \min\{V_l | i \in D_l\}$, $i = 1, \dots, n$. We assume that V_1, \dots, V_{n+p} are absolutely continuously distributed, independent random variables with distribution functions $F_l(t)$, survival functions $\bar{F}_l(t)$, densities $f_l(t)$ and failure rates $\lambda_l(t)$, $l = 1, 2, \dots, n + p$. Due to the absolute continuity of the V_l 's a subset $A \subset E$ satisfies $P(D = A) > 0$ if and only if A is a cut set, i.e. $\phi(x_1, \dots, x_n) = 0$ if $x_i = 0$ for $i \in A$, and there exists a shock $j \in S$ such that $D_j \subset A$ and $A - D_j$ is not a cut set. Such a set A is called a fatal set, and we introduce $\mathcal{A} = \{A_1, \dots, A_m\}$, the set of fatal sets. For $i \in \{1, 2, \dots, m\}$ define $G_i(t) = P(T \leq t, D = A_i)$. Then the likelihood function for the data $(T = t, D = A_i)$ is $g_i(t) = G'_i(t)$, which is the Radon-Nikodym derivative of the probability measure for (T, D) with respect to the Lebesgue measure \times counting measure on $[0, \infty) \times \{1, 2, \dots, m\}$. Gåsemyr & Natvig (1995b) gives a procedure for deriving a computationally efficient expression for the likelihood, see

their (2.5). In the present paper, we will only be interested in the general form of the likelihood, which can be easily derived. For $i \in \{1, \dots, m\}$, define

$$\mathcal{B}_i = \{(B, j) \text{ with } B \subset S, j \in S \mid D_{B \cup \{j\}} = A_i, D_B \text{ is not a cut set}\}$$

The likelihood can then be written as

$$g_i(t) = \sum_{(B, j) \in \mathcal{B}_i} \prod_{l \in B} F_l(t) \prod_{l \in S-B} \bar{F}_l(t) \lambda_j(t) \quad (2.1)$$

Specializing to the exponential case so that the joint distribution of T_1, T_2, \dots, T_n is a multivariate exponential distribution of the Marshall-Olkin type, see Marshall & Olkin (1967), we get in particular the following lemma.

Lemma 1 *Suppose that in the shock model the times to shocks V_l are exponentially distributed with failure rates $\theta_l, l = 1, 2, \dots, n + p$. Then the likelihood for the autopsy data $(T = t, D = A_i)$ can be written in the form*

$$g_i(t) = \sum_{k=1}^K \prod_{l \in B_k} h(\theta_l; 1, 0, t) \prod_{l \in C_k} h(\theta_l; 1, t) h(\theta_{j_k}; 2, t), \quad (2.2)$$

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

Proof: One version of equation (2.2) follows by indexing the members of the pairs (B, j) of \mathcal{B}_i in (2.1) by k and putting $C_k = S - (B_k \cup \{j_k\})$. The more general form (2.2) where we can have $B_k \cup C_k \cup \{j_k\} \neq S$, may, however, arise by applying the procedure for calculating the likelihood given in Gåsemeyr & Natvig (1995b), see their (2.11).

The following result is also given in Gåsemeyr & Natvig (1995b).

Theorem 1

a) *Suppose that the rates $\theta_1, \theta_2, \dots, \theta_{n+p}$ for the occurrence of shocks in an exponential shock model for a binary monotone system (E, ϕ) have a joint prior distribution of the form*

$$\pi(\underline{\theta}) \propto \sum_{j=1}^J \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) = \sum_{j=1}^J \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l}, \underline{t}_{j,l})^{-1} g(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) \quad (2.3)$$

Then the posterior distribution of $\underline{\theta}$ given the autopsy data $(T = t, D = A)$ with likelihood function given by (2.2) is of the form

$$\begin{aligned} \pi(\underline{\theta} | T = t, D = A) &\propto \sum_{j=1}^J \sum_{k=1}^K \left[\prod_{l \in B_k} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}, t) \right. \\ &\quad \left. \prod_{l \in C_k} h(\theta_l; a_{j,l}, b_{j,l} + t, \underline{t}_{j,l}) \right] h(\theta_{j_k}; a_{j,j_k} + 1, b_{j,j_k} + t, \underline{t}_{j,j_k}) \\ &\quad \times \left[\prod_{l \in S - (B_k \cup C_k \cup \{j_k\})} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) \right] \end{aligned} \quad (2.4)$$

b) *The class of distributions of the form (2.3) is a natural conjugate class of priors for the exponential autopsy shock model.*

c) *Suppose the prior distribution*

$$\pi(\underline{\theta}) = \prod_{l=1}^{n+p} g(\theta_l; a_l, b_l), \quad (2.5)$$

for $\underline{\theta}$ is updated with autopsy data from r independent systems, of the form $(T_1 = t_1, D_1 = A_{i_1}, \dots, T_r = t_r, D_r = A_{i_r})$. Then the posterior distribution is of the form (2.3), with $\underline{t}_{j,l}$ a subvector of $\underline{t} = (t_1, t_2, \dots, t_r)$ for all $j = 1, \dots, J, l = 1, \dots, n + p$.

Proof: a) is a straightforward application of Bayes theorem. b) follows since (2.4) is of the same general form as (2.3). c) follows by repeated use of a).

Distributions of the form (2.3) arise naturally in the model if one chooses a prior of the form (2.5), or if information used to construct a prior can be thought of as consistent with autopsy data. A wider class of priors is obtained if one allows for a positive weight w_j for the j th summand in (2.3). An even wider class of natural conjugate priors is the class of distributions of the form

$$\pi(\underline{\theta}) = \sum_{j=1}^J w_j \prod_{l=1}^{n+p} g(\theta_l; a_{j,l}, b_{j,l}), \quad (2.6)$$

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 $\pi(\underline{\theta}) \geq 0$ for all vectors $\underline{\theta}$ with positive entries. Posterior densities of the form (2.6) arise, like distributions of the form (1.6) in the single component case, if the factors $(1 - e^{-\theta_l t}), l \in B_k, k = 1, \dots, K$ appearing in (2.2), are multiplied out. This may potentially increase the number of summands in the likelihood from K to $\sum_{k=1}^K 2^{|B_k|}$.

Admittedly, when updating with autopsy data, both complexity of the system and abundance of data may lead to prohibitively complicated posterior distributions. We do, however, primarily have in mind situations where data are scarce, which is typical in reliability. However, the following discussion strongly indicates that expressing the posterior distribution in terms of the generalized gamma distributions as in (2.3) leads to considerably more computationally tractable expressions than using only ordinary gamma distributions, as in (2.6).

Referring to (2.3), let $\underline{t} = (t_1, t_2, \dots, t_r)$ be a vector with positive entries such that $\underline{t}_{j,l}$ is a subvector of \underline{t} for all $j = 1, \dots, J, l = 1, \dots, n + p$ (cf. c) of Theorem 1). For $j = 1, \dots, J, l = 1, \dots, n + p$ define $\underline{d}_{j,l} = (d_{j,l,1}, d_{j,l,2}, \dots, d_{j,l,r})$ by putting $d_{j,l,i} = 1$ if t_i occurs as an entry in the subvector $\underline{t}_{j,l}$ and 0 otherwise.

Defining $\underline{t}(\underline{d}_{j,l})$ = the vector whose entries are the non-zero entries of $(t_1 d_{j,l,1}, \dots, t_r d_{j,l,r})$ recovers $\underline{t}_{j,l}$ from $\underline{d}_{j,l}$. Note that the normalizing constant corresponding to the l th factor of the j th summand of (2.3) can be calculated as in (1.2)

$$\gamma(a_{j,l}, b_{j,l}, \underline{t}_{j,l}) = (\Gamma(a_{j,l}))^{-1} \left(\sum_{\underline{d} \in \{0,1\}^r | \underline{d} \leq \underline{d}_{j,l}} (b_{j,l} + \underline{d} \cdot \underline{t})^{-a_{j,l}} \right)^{-1} \quad (2.7)$$

Introduce $\underline{d}_j = (d_{j,1,1}, \dots, d_{j,1,r}, \dots, d_{j,n+p,1}, \dots, d_{j,n+p,r})$, i.e. the vector made up by the subvectors $\underline{d}_{j,l}$, $l = 1, \dots, n+p$. The distribution in (2.3) can then be written in the form (2.6) as

$$\begin{aligned} \pi(\underline{\theta}) &\propto \sum_{j=1}^J \prod_{l=1}^{n+p} \prod_{i=1}^r (1 - e^{-\theta_l t_i})^{d_{j,l,i}} h(\theta_l; a_{j,l}, b_{j,l}) \\ &= \sum_{j=1}^J \sum_{\underline{c} \in \{0,1\}^{r(n+p)} | \underline{c} \leq \underline{d}_j} (-1)^{|\underline{c}|} \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l} + \sum_{i=1}^r c_{(l-1)r+i} t_i) \end{aligned} \quad (2.8)$$

It is worth noting that the notation introduced in this discussion may be conveniently used when updating π with data. The updated distribution (2.4) is determined by the binary vectors $\underline{d}_{(j-1)K+k,l} = (\underline{d}_{j,l}, I(l \in B_k))$ relating to the vector (\underline{t}, t) , together with the parameters $a_{(j-1)K+k,l} = a_{j,l} + I(l = j_k)$ and $b_{(j-1)K+k,l} = b_{j,l} + tI(l \in C_k)$, $j = 1, \dots, J$, $k = 1, \dots, K$, $l = 1, \dots, n+p$.

To calculate the distribution from (2.8), one must calculate a weight for each of the $\sum_{j=1}^J 2^{|\underline{d}_j|}$ summands, each of which involves a product of $n+p$ factors of the form $\Gamma(a)(b + \underline{d} \cdot \underline{t})^{-a}$. In contrast, (2.3) has only J summands, and the weight corresponding to the j th summand involves $\sum_{l=1}^{n+p} 2^{|\underline{d}_{j,l}|}$ terms of the form $\Gamma(a)(b + \underline{d} \cdot \underline{t})^{-a}$ (cf. (2.7)). Note that

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

Hence,

$$\sum_{l=1}^{n+p} 2^{|\underline{d}_{j,l}|} \leq 2^{|\underline{d}_j|} \quad (2.9)$$

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 (2.3) is considerably more efficient computationally than (2.6) in the present model, thus providing a good case for the usefulness of the generalized gamma distribution.

We conclude this section by generalizing (1.4) and (1.5) to the shock model, giving expressions also presented in Gåsemyr & Natvig (1995b). Estimation of θ_m , $m = 1, \dots, n + p$ in analogy with (1.4) leads to the following straightforward calculation by applying (2.3).

$$\begin{aligned}
E(\theta_m) &= \frac{\int_{[0, \infty]^{n+p}} \theta_m \sum_{j=1}^J \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) d\theta_1 \cdots d\theta_{n+p}}{\int_{[0, \infty]^{n+p}} \sum_{j=1}^J \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) d\theta_1 \cdots d\theta_{n+p}} \\
&= \frac{\sum_{j=1}^J \prod_{l=1}^{n+p} \gamma(a_{j,l} + I(l = m), b_{j,l}, \underline{t}_{j,l})^{-1}}{\sum_{j=1}^J \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l}, \underline{t}_{j,l})^{-1}}
\end{aligned} \tag{2.10}$$

We now turn to the computation of the predictive survival probability $P(T > t)$ based on a distribution π of the form (2.3). A similar problem has been considered in Gåsemyr & Natvig (1995a) where $P(T > t)$ is updated with life data from identical systems. This follows the paradigm of Natvig & Eide (1987) of using the distribution of the component reliabilities, possibly updated with data at the component level, as a basis for computing the distribution of system reliability, and then updating this distribution with data at the system level. In practice, however, data from operation of systems may contain information about specific components, as is the case with autopsy data. Then one cannot separate the component and system level as prescribed by Natvig & Eide (1987). In our situation we assume that the autopsy data has already been incorporated in the distribution π , as described above.

Conditionally on $\underline{\theta}$, the system survival probability can be written as

$$P(T > t | \underline{\theta}) = \sum_{A \subseteq E} \delta_A \prod_{k \in E_A} e^{-\theta_k t}, \tag{2.11}$$

where δ is the signed domination function of (E, ϕ) defined through the equation

$$\phi(\underline{x}) = \sum_{A \subseteq E} \delta_A \prod_{k \in A} x_k,$$

see Gåsemyr & Natvig (1995a). Combining (2.3) and (2.11) we then obtain

$$\begin{aligned}
P(T > t) &= \int_{[0, \infty]^{n+p}} P(T > t | \underline{\theta}) \pi(\underline{\theta}) d\underline{\theta} \\
&= \frac{\int_{[0, \infty]^{n+p}} \sum_{A \subseteq E} \sum_{j=1}^J \delta_A \prod_{k \in E_A} e^{-\theta_k t} \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) d\theta_1 \cdots d\theta_{n+p}}{\int_{[0, \infty]^{n+p}} \sum_{j=1}^J \prod_{l=1}^{n+p} h(\theta_l; a_{j,l}, b_{j,l}, \underline{t}_{j,l}) d\theta_1 \cdots d\theta_{n+p}} \\
&= \frac{\sum_{A \subseteq E} \sum_{j=1}^J \delta_A \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l} + I(l \in E_A) t, \underline{t}_{j,l})^{-1}}{\sum_{j=1}^J \prod_{l=1}^{n+p} \gamma(a_{j,l}, b_{j,l}, \underline{t}_{j,l})^{-1}}
\end{aligned} \tag{2.12}$$

3 Simulation procedures for parameter estimation and prediction

Let us start by returning to Example 1 in Section 1. Remember that the prior distribution of θ was assumed to be gamma with parameters a, b , and that we ended up with the posterior distribution being the generalized gamma distribution $g(\theta, a, b + \underline{d} \cdot \underline{t}, t_{i_1}, \dots, t_{i_m})$.

Here $m = \sum_{i=1}^r (1 - d_i)$ is the number of left censored observations. From the expressions for $E(\theta)$ and the predictive survival probability, $P(T > t)$, given by respectively (1.4) and (1.5), we see that for both of them the number of summands both in the numerator and denominator is 2^m . The same is true in Example 2. If m is large, the computational complexity can be an obstacle and simulation may be an alternative. Especially, importance sampling seems relevant, see Geweke (1989).

Moving to our autopsy model in Section 2 the corresponding expressions for $E(\theta_l), l = 1, \dots, n + p$ and $P(T > t)$ are given by respectively (2.10) and (2.12). The number of summands both in the numerator and denominator of (2.10) and in the denominator of (2.12) is J , whereas in the numerator of (2.12) the number is $|\{A | \delta_A \neq 0\}| J \leq 2^n J$. Here, n is the number of components. Furthermore, if $\pi(\underline{\theta})$ given by (2.3), entering in (2.10) and (2.12), 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 $\gamma(a_{j,l}, b_{j,l}, \underline{t}_{j,l})^{-1}$, see (2.7), increases with a factor of 2 for each new observation, i.e. with a factor of 2^r altogether. Thus the computational complexity of (2.12) may be formidable. In some cases it may even be impossible to calculate the signed domination function δ . To overcome such problems simulation can be the only alternative.

Again importance sampling may be relevant. However, we will in addition suggest some simulation procedures based on the Metropolis-Hastings algorithm, see Smith & Roberts

(1993), Tierney (1994), Besag et al. (1995) and Chib & Greenberg (1995) with corresponding discussions and references. We assume that it is easy to simulate from π_0 . π_0 may for instance be a product of independent gamma distributions. We denote by $L(\underline{\theta}|t_i, A_{j_i})$ the likelihood for $\underline{\theta}$ given the autopsy data ($T_i = t_i, D_i = A_{j_i}$). Thus, we have by Bayes theorem

$$\pi(\underline{\theta}) = \pi(\underline{\theta}|T_1 = t_1, D_1 = A_{j_1}, \dots, T_r = t_r, D_r = A_{j_r}) \propto \pi_0(\underline{\theta}) \prod_{i=1}^r L(\underline{\theta}|t_i, A_{j_i}) \quad (3.1)$$

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

$$\beta(\underline{\theta}_k, \underline{\theta}') = (\pi(\underline{\theta}')\pi_0(\underline{\theta}_k))/(\pi(\underline{\theta}_k)\pi_0(\underline{\theta}')) = \left(\prod_{i=1}^r L(\underline{\theta}'|t_i, A_{j_i})\right)/\left(\prod_{i=1}^r L(\underline{\theta}_k|t_i, A_{j_i})\right) \quad (3.2)$$

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

$$N^{-1} \sum_{k=1}^N P(T > t|\underline{\theta}_k), \quad (3.3)$$

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 (3.3) may be obtained by replacing the summands of (3.3) by approximate values based for instance on the bounds for the reliability of a shock system given in Gåsemeyr & Natvig (1995a).

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 $\pi(\cdot|T_1 = t_1, D_1 = A_{j_1}, \dots, T_{r+1} = t_{r+1}, D_{r+1} = A_{j_{r+1}})$ to be faster if drawing candidate values $\underline{\theta}'$ from π_0 could be replaced by drawing from $\pi(\cdot|T_1 = t_1, D_1 = A_{j_1}, \dots, T_r = t_r, D_r = A_{j_r})$. It may therefore be profitable to draw from an easily simulated approximation $\pi^{(r)}$ to the latter distribution. One possible choice for $\pi^{(r)}$ is a product of gamma distributions with the correct marginal expectations, see (2.10), and correct variances. 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 $E(\theta_l|T_1 = t_1, D_1 = A_{j_1}, \dots, T_r = t_r, D_r = A_{j_r})$ and $E(\theta_l^2|T_1 = t_1, D_1 = A_{j_1}, \dots, T_r = t_r, D_r = A_{j_r})$, $l = 1, \dots, n + p$, may be estimated by replacing $P(T > t|\underline{\theta}_k)$ with $\theta_{k,l}$ and $\theta_{k,l}^2$ respectively in (3.3). The original simulation procedure would then be modified by modifying $\beta(\underline{\theta}_k, \underline{\theta}')$ to

$$\beta(\underline{\theta}_k, \underline{\theta}') = (\pi_0(\underline{\theta}') \prod_{i=1}^{r+1} L(\underline{\theta}'|t_i, A_{j_i}) \pi^{(r)}(\underline{\theta}_k))/(\pi_0(\underline{\theta}_k) \prod_{i=1}^{r+1} L(\underline{\theta}_k|t_i, A_{j_i}) \pi^{(r)}(\underline{\theta}')) \quad (3.4)$$

The effect on the convergence rate of choosing a prior and/or a distribution for candidate values $\underline{\theta}'$ 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 approach described above, is quite general. Here it is used once for each data set, the starting point being the arrival of additional data. It should be mentioned, as also has been pointed out to us by Arnaldo Frigessi, that the approach could also be used successively for the same data set. This gives rise to questions on stopping rules, convergence properties etc. and lots of space for computational experiments.

Another suggestion of an “adaptive sampler” is to recursively use the output of the simulations based on the data $(T_1 = t_1, D_1 = A_{j_1}, \dots, T_r = t_r, D_r = A_{j_r})$ as input simulations when updating with the new data $(T_{r+1} = t_{r+1}, D_{r+1} = A_{j_{r+1}})$, $r = 1, 2, \dots$. The following calculations show that this all the way simulates Markov chains with the desired stationary distributions.

It should be noted that this idea may be viewed as basically the same resampling idea as presented in Smith & Gelfand (1992) whereby samples from one distribution may be modified to form samples from another distribution. The two methods given in the latter paper are the Rejection Method and the Weighted Bootstrap. In Rubin (1988) the latter is referred to as SIR (sampling importance resampling). This is applied successfully in Gordon et al. (1993) to nonlinear/non-Gaussian Bayesian forecasting.

The first step is identical to the one leading to (3.2). Hence, with slightly extended notation

$$\begin{aligned}\beta_1(\underline{\theta}, \underline{\theta}') &= L(\underline{\theta}' | t_1, A_{j_1}) / L(\underline{\theta} | t_1, A_{j_1}) \\ \alpha_1(\underline{\theta}, \underline{\theta}') &= \min\{1, \beta_1(\underline{\theta}, \underline{\theta}')\}\end{aligned}\tag{3.5}$$

We also introduce the transition probabilities

$$p_1(\underline{\theta}, \underline{\theta}') = \pi_0(\underline{\theta}') \alpha_1(\underline{\theta}, \underline{\theta}') \stackrel{\text{def}}{=} q_2(\underline{\theta}, \underline{\theta}'),\tag{3.6}$$

corresponding to a Markov chain with stationary distribution $\pi(\underline{\theta} | T_1 = t_1, D_1 = A_{j_1})$.

To obtain a Markov chain with stationary distribution $\pi(\underline{\theta} | T_1 = t_1, D_1 = A_{j_1}, T_2 = t_2, D_2 = A_{j_2})$, following the general Metropolis-Hastings algorithm, introduce

$$\begin{aligned}
\beta_2(\underline{\theta}, \underline{\theta}') &= \frac{\pi(\underline{\theta}'|t_1, A_{j_1}, t_2, A_{j_2})q_2(\underline{\theta}', \underline{\theta})}{\pi(\underline{\theta}|t_1, A_{j_1}, t_2, A_{j_2})q_2(\underline{\theta}, \underline{\theta}')} \\
&= \frac{\pi_0(\underline{\theta}') \prod_{i=1}^2 L(\underline{\theta}'|t_i, A_{j_i})\pi_0(\underline{\theta}) \min\{1, L(\underline{\theta}|t_1, A_{j_1})/L(\underline{\theta}'|t_1, A_{j_1})\}}{\pi_0(\underline{\theta}) \prod_{i=1}^2 L(\underline{\theta}|t_i, A_{j_i})\pi_0(\underline{\theta}') \min\{1, L(\underline{\theta}'|t_1, A_{j_1})/L(\underline{\theta}|t_1, A_{j_1})\}} \quad (3.7) \\
&= L(\underline{\theta}'|t_2, A_{j_2})/L(\underline{\theta}|t_2, A_{j_2}),
\end{aligned}$$

the second equality following from (3.5) and (3.6). Also introduce

$$\alpha_2(\underline{\theta}, \underline{\theta}') = \min\{1, \beta_2(\underline{\theta}, \underline{\theta}')\}, \quad (3.8)$$

giving the transition probabilities of the desired Markov chain

$$\begin{aligned}
p_2(\underline{\theta}, \underline{\theta}') &= q_2(\underline{\theta}, \underline{\theta}')\alpha_2(\underline{\theta}, \underline{\theta}') \\
&= \pi_0(\underline{\theta}') \min\{1, L(\underline{\theta}'|t_1, A_{j_1})/L(\underline{\theta}|t_1, A_{j_1})\} \min\{1, L(\underline{\theta}'|t_2, A_{j_2})/L(\underline{\theta}|t_2, A_{j_2})\} \quad (3.9)
\end{aligned}$$

$$\stackrel{\text{def}}{=} q_3(\underline{\theta}, \underline{\theta}'),$$

having applied (3.5)-(3.8).

Repeating this procedure, all the way we establish the correct transition probabilities and hence Markov chains with the correct stationary distributions. Note especially that

$$\beta_{r+1}(\underline{\theta}, \underline{\theta}') = L(\underline{\theta}'|t_{r+1}, A_{j_{r+1}})/L(\underline{\theta}|t_{r+1}, A_{j_{r+1}}), \quad r = 0, 1, \dots, \quad (3.10)$$

the simplicity of which, for instance compared to (3.4), suggests that this “adaptive sampler” may be computationally efficient.

Note that in this approach at stationarity, the input variables for the $(r+1)$ th run, being the output variables for the r th run, possess the desired joint distribution. In particular, the variables do not only have the correct marginals, but also the correct dependence structure, which may be a great advantage. In contrast, in our first approach, the marginals are supposedly good approximations, but the components of $\underline{\theta}$ in the proposal distribution are always independent (except in the suggestion of using convex combinations of generalized gamma distributions).

Thorough simulations are planned to investigate the adequacy of the suggestions in this section.

Acknowledgement

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

References

- Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statist. Science*, **10**, 3–66.
- Boyles, R.A. & Samaniego, F.J. (1984). Modelling and inference for multivariate binary data with positive dependence. *J. Amer. Statist. Ass.* **79**, 188–193.
- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *Amer. Statist.* **49**, 327–335.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57**, 1317–1339.
- Gordon, N.J., Salmond, D.J. & Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140**, 107–113.
- Gåsemyr, J. & Natvig, B. (1995a). Some aspects of reliability analysis in shock models. *Scand. J. Statist.* **22**, 385–393.
- Gåsemyr, J. & Natvig, B. (1995b). The posterior distribution of the parameters of component lifetimes based on autopsy data in a shock model. Submitted to *Scand. J. Statist.*
- Marshall, A.W. & Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Ass.* **62**, 30–44.
- Meilijson, I. (1981). Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine. *J. Appl. Prob.* **18**, 829–838.
- Natvig, B. & Eide, H. (1987). Bayesian estimation of system reliability. *Scand. J. Statist.* **14**, 319–327.
- Rubin, D.B. (1988). Using the SIR algorithm to simulate posterior distributions. In *Bayesian Statistics 3*, eds. Bernardo, J.M., DeGroot, M.H., Lindley, D.V. & Smith A.F.M. Oxford University Press, 395–402.
- Smith, A.F.M. & Gelfand, A.E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *Amer. Statist.* **46**, 84–88.
- Smith, A.F.M. & Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and

related Markov Chain Monte Carlo methods. J.R. Statist. Soc. B. **55**, 3–23, 53–102.

Tierney, L. (1994). Markov chains for exploring posterior distributions. Ann. Statist. **22**, 1701–1762.