# Bayesian analysis of quasi-life tables

David A. Stephens
d.stephens@imperial.ac.uk

Martin J. Crowder
m.crowder@imperial.ac.uk
*Department of Mathematics , Imperial College London, 180 Queen's Gate, London SW7 2BZ, England*

September 22, 2005

**Abstract**. Quasi-life tables, in which the data arise from many concurrent, independent, discrete-time renewal processes, were defined by Baxter (1994), who outlined some methods for estimation. The processes are not observed individually, only the total numbers of renewals at each time point are observed. Crowder and Stephens (2003) implemented a formal estimating-equation approach that invokes large-sample theory. However, these asymptotic methods fail to yield sensible estimates for smaller samples. In this paper, we implement a Bayesian analysis based on MCMC computation that works equally well for large and small sample sizes. We give three simulated examples, studying the Bayesian results, the impact of changing prior specification, and empirical properties of the Bayesian estimators of the lifetime distribution parameters. We also study the Baxter (1994) data, and uncover structure that has not been commented upon previously.

**Keywords**: Discrete-time renewal process, quasi-life tables, recurrent events data, survival analysis, Markov chain Monte Carlo, missing-data likelihood

## 1 Introduction

Repairable systems, and the stochastic processes that underpin them, are topics of considerable interest in reliability, econometrics, finance and many other fields of application. For example, Baxter (1994) studied a situation in which a manufactured unit contains a replaceable component whose failure is fatal to the unit. A component is installed at time 0 and is liable to failure at subsequent times 1, 2, ...; if it fails at time $s$, a new component is installed, and the new component is liable to failure at times $s+1, s+2, \ldots$. Some of the failure data presented in Baxter (1994) are given in Table 1, which is to be read across rows: these are the aggregated numbers of failures per month over 58 months with 4145 new units installed each month. For instance, there are no failures until month 8, by which time $8 \times 4145$ units are in operation (these having been installed at months 0,...,7). Such data constitute an example of

Table 1: Failure data from Baxter (1994, Table 1)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 6 | 6 | 1 | 11 |
| 26 | 18 | 22 | 28 | 19 | 18 | 48 | 35 | 46 | 69 | 48 | 73 |
| 141 | 116 | 96 | 120 | 99 | 110 | 140 | 97 | 126 | 99 | 57 | 90 |
| 132 | 98 | 113 | 178 | 115 | 99 | 227 | 139 | 185 | 169 | 226 | 230 |
| 262 | 109 | 90 | 224 | 207 | 202 | 269 | 149 | 163 | 259 | | |

a discrete-time data-collection exercise. In the following subsections we describe suitable statistical models for the analysis of such data.

The complicating feature arising in the analysis is that only monthly-aggregate data are available, that is, we do not observe individual failure and repair events, but rather only totals of numbers of repairs across a cohort of individual experimental units. This complication makes direct statistical inference difficult to implement.

In this paper, we consider a Bayesian solution to the inference problem. We feel that the Bayesian solution offers a full representation of uncertainty, and also facilitates coherent means for prediction. However, some alternative approaches are available. For example, Baxter (1994) used a likelihood approach based on the basic renewal properties of the recorded data. Maximum likelihood estimates of the lifetime distribution probabilities may also be obtained using an EM approach; see, for example, Karim et al. (2001) for a successful application in a related data analysis problem. Such an approach may be feasible here, but is not the focus of this paper. Finally, Crowder and Stephens (2003) developed estimating equations for this type of data that rely on large-sample theory for their theoretical properties. Unfortunately, this approach can fail when the data are not extensive; we demonstrate this in section 8.1.

## 2 A discrete-time failure model

In the basic scenario a component is installed at time $r$, and is liable to failure at time $r+s$ $(s=1,2\ldots)$. When it fails it is immediately repaired or replaced, and so on. At any given time there will be at most one failure incident, and a typical repair record for the component might look like the following:

| Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Repairs | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |

This record corresponds to failures at times 3, 5, 9, and 10, giving lifetimes of lengths 3, 2, 4, 1 followed by a lifetime censored at two months.

To construct a general model for this experimental situation, a discrete lifetime distribution is considered: this will be denoted by $\boldsymbol{q}=(q_1,q_2,\ldots)$, where

$$q_s = \text{pr}(\text{component fails at age } s) \quad (s=1,2,\ldots).$$

We will refer to this specification as non-parametric, though it could be said that the $q_s$ are the parameters; subsequent modelling may use a restricted parametric form for $\boldsymbol{q}$. In this paper we concentrate in the main on the non-parametric case. We will also use the survivor function,

$$Q(s)=\text{pr}(X>s)=1-q_1-q_2-\ldots-q_s \quad (s=1,2,\ldots),$$

with $Q(0)=1$, and the discrete hazards,

$$h_s=\text{pr}(X=s)/\text{pr}(X\geq s)=q_s/Q(s-1) \quad (s=1,2,\ldots).$$

We note for reference the standard relationships

$$q_s=h_s\prod_{j=1}^{s-1}(1-h_j), \qquad Q(s)=\prod_{j=1}^{s}(1-h_j) \quad (s=1,2,\ldots) \tag{2.1}$$

## 3 Aggregation of renewal processes

As identified above, data aggregation complicates the inference procedure. We consider the discrete-time renewal process defined in Section 2. Suppose that $b_r$ units are installed at time $r$ $(r=0,1,2,\ldots)$ and operate independently. The failure record up to time $t$ for a unit among the cohort installed at time $r$ $(r<t)$ can be represented as a binary vector of length $t-r$, with 1 for a component failure and 0 otherwise. The record for the $i$th unit in the $r$th cohort is then

$$\boldsymbol{c}_i^{(r)}=(c_{i,r+1}^{(r)},c_{i,r+2}^{(r)},\ldots,c_{it}^{(r)}),$$

where $c_{i,r+s}^{(r)}=1$ if the unit has a failure at time $r+s$, and $c_{i,r+s}^{(r)}=0$ otherwise. Schematically, the parallel discrete-time renewal processes for the $b_r$ units can be displayed as in Table 2, in which the $i$th row is $\boldsymbol{c}_i^{(r)}$ and

$$d_{r,r+s}=\sum_{i=1}^{b_r}c_{i,r+s}^{(r)}$$

is the total number of failures in cohort $r$ at time $r+s$. In practice, the $c_{i,r+s}^{(r)}$ are often not individually recorded,

Table 2: Array of failure indicators for cohort of units installed at time $r$

| time | $r+1$ | $r+2$ | $\ldots$ | $t$ |
|---|---|---|---|---|
| unit $i=1$ | $c_{1,r+1}^{(r)}$ | $c_{1,r+2}^{(r)}$ | $\ldots$ | $c_{1t}^{(r)}$ |
| 2 | $c_{2,r+1}^{(r)}$ | $c_{2,r+2}^{(r)}$ | $\ldots$ | $c_{2t}^{(r)}$ |
| . | . | . | $\ldots$ | . |
| $b_r$ | $c_{b_r,r+1}^{(r)}$ | $c_{b_r,r+2}^{(r)}$ | $\ldots$ | $c_{b_rt}^{(r)}$ |
| Total | $d_{r,r+1}$ | $d_{r,r+2}$ | $\ldots$ | $d_{rt}$ |

the available data comprising only the totals, $d_{r,r+s}$. Sometimes the $d_{r,r+s}$ themselves are not individually recorded, but only the totals over cohorts. This latter case occurs when the manufacturer keeps a record only of the numbers of replacement components supplied to the customers at times $1,2,\ldots$.

In subsequent sections, we will consider a likelihood-based approach based on the non-parametric model of the failure distribution, and a Bayesian solution to the inference problem. The Bayesian approach enables full inference about the failure distribution, and predictions about future failures, but involves complicated numerical computation of the posterior. We introduce a Markov chain Monte Carlo (MCMC) scheme that facilitates the computation of the missing-data likelihood and the posterior distribution for the failure distribution, and yields a straightforward method of prediction. We also address some aspects of prior specification. Finally, we illustrate the computational method with examples, concluding with an analysis of the Baxter (1994) data.

# 4 Likelihood function

In this and subsequent sections, we consider a likelihood-based approach that eventually leads to a Bayesian solution via MCMC. We first consider the construction of a tractable likelihood function by using a data-augmentation approach.

## *4.1 Constructing the likelihood function*

Suppose that unit $i$, installed at time $r$, suffers $m_i$ component-failures up to time $t$. The completed lifetimes of the failed components (i.e. the lengths of the intervals between 1s in the $c_i^{(r)}$-sequence) are $l_{i1}, \ldots, l_{im_i}$, say, and $l_{i,m_i+1}$ is the right-censored lifetime of the component in place and still functioning at time $t$; $l_{i,m_i+1} = 0$ if the $m_i$th failure is actually at time $t$. With these definitions, the probability for the failure record of unit $i$ is

$$p(c_i^{(r)}|\boldsymbol{q}) = \prod_{j=1}^{m_i} q_{l_{ij}} \times Q(l_{i,m_i+1}) \tag{4.1}$$

and the joint probability for all $b_r$ records of the $r$-cohort, i.e. for the whole $b_r \times (t - r + 1)$ table, is

$$p(c_1^{(r)}, \ldots, c_{b_r}^{(r)}|\boldsymbol{q}) = \prod_{i=1}^{b_r} p(c_i^{(r)} \mid \boldsymbol{q}). \tag{4.2}$$

This likelihood cannot be written effectively as a function of the column totals $d_{rs}$ alone because the $d_{rs}$ are not, in general, sufficient statistics. To see this, consider the two tables

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

which have the same column totals but different probabilities, $q_1(1 - q_1)q_2$ and $q_1^2(1 - q_1 - q_2)$, respectively. The likelihood function based on the $d_{rs}$ alone is obtained by summing (4.2) over all tables with rows $\boldsymbol{c}_i^{(r)}$ generated by repositioning the 0s and 1s in each column so that the $d_{rs}$-values are preserved: there are $\binom{b_r}{d_{rs}}$ different arrangements in column $s$, and the total number of contributing tables is the product of these binomial coefficients over $s = r+1, \ldots, t$.

The failure records can be assembled in a $b_+ \times t$ table, say $C$, where $b_+ = \sum_{r=0}^{t-1} b_r$ and $C$ is structured as $C = (C_1^{\mathrm{T}}, \ldots, C_t^{\mathrm{T}})^{\mathrm{T}}$, in which the $i$th row of $C_{r+1}$ contains the repair record $c_i^{(r)}$ of component $i$ in cohort $r$. Thus, $C_{r+1}$ $(r = 0, \ldots, t-1)$ is a $b_r \times t$ matrix of 0s and 1s:

$$C_{r+1} = \begin{bmatrix} 0 & \cdots & 0 & c_{1,r+1}^{(r)} & c_{1,r+2}^{(r)} & \cdots & c_{1t}^{(r)} \\ 0 & \cdots & 0 & c_{2,r+1}^{(r)} & c_{2,r+2}^{(r)} & & c_{2t}^{(r)} \\ \vdots & & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & c_{b_r,r+1}^{(r)} & c_{b_r,r+2}^{(r)} & & c_{b_r t}^{(r)} \end{bmatrix},$$

in which the first $r$ columns contain only zeros. From (4.1), the log-likelihood function for $c_i^{(r)}$, the $i$th row of $C_r$, is

$$\prod_{j=1}^{m_i^{(r)}} \{q_{l_{ij}^{(r)}} Q(l_{i,m_i^{(r)}+1}^{(r)})\}, \tag{4.3}$$

in terms of the lifetime lengths $l_{ij}^{(r)}$ and the censored final lifetime $l_{i,m_i^{(r)}+1}^{(r)}$, and where $m_i^{(r)}$ is the number of failures of the $i$th component in the $r$th cohort. The log-likelihood for $C_r$ is obtained by summing (4.3) over $i = 1, \ldots, b_r$, and then that for $C$ is obtained by summing again over $r = 0, \ldots, t-1$. The constraints on $C$ are that the column sums are $d_{+s} = \sum_{r=0}^{s-1} d_{rs}$. Thus, with $b_{+s} = \sum_{r=0}^{s} b_r$, there are $\binom{b_{+s}}{d_{+s}}$ possible configurations of 0s and 1s in column $s$ of $C$. For the data in Baxter (1994, Table 1) $b_r = 4145$ and $t = 58$, and only the $d_{+s}$ were recorded. The number of possible tables consistent with these totals is in excess of $10^{6870}$ so it is clearly not feasible to marginalize by summing over all tables for inference about $\boldsymbol{q}$. An approximate likelihood inference method is described in the next section.

### 4.2 A Monte Carlo Approximation to the Likelihood

For likelihood inference, or for use in a Bayesian analysis, it is possible to construct a Monte Carlo approximation that avoids the need for enumeration of the likelihood function for the potentially vast number of missing data tables. Let $d = (d_{+1}, \ldots, d_{+t})$, and denote the space of all possible failure record profiles as $\mathbb{C}$, and its restriction to those profiles compatible with row-total data $\boldsymbol{d}$ as $\mathbb{C}_d$. Then the likelihood $p(\boldsymbol{d} \mid \boldsymbol{q})$ for probabilities $\boldsymbol{q}$ can be written

$$p(\boldsymbol{d} \mid \boldsymbol{q}) = \sum_{C \in \mathbb{C}_d} p(\boldsymbol{d}, C \mid \boldsymbol{q}) = \sum_{C \in \mathbb{C}_d} p(C \mid \boldsymbol{q}) \tag{4.4}$$

as $C$ implies $\boldsymbol{d}$, where $p(C \mid \boldsymbol{q})$ is given by (4.3) summed over all possible tables that satisfy the column total constraints. The exhaustive summation over the set $\mathbb{C}_d$ is prohibitive, even for relatively small $t$ and $b$. Let $n_C = |\mathbb{C}_d|$ denote the cardinality of $\mathbb{C}_d$, that is, the number of tables that satisfy the column constraints. Then

$$p(\boldsymbol{d} \mid \boldsymbol{q}) = \sum_{C \in \mathbb{C}_d} p(C \mid \boldsymbol{q}) = n_C \sum_{i=1}^{n_C} p(C^{(i)} \mid \boldsymbol{q}) \left( \frac{1}{n_C} \right) = n_C E\left[p(C \mid \boldsymbol{q})\right] \propto E\left[p(C \mid \boldsymbol{q})\right]$$

where the expectation is taken with respect to a uniform distribution on the elements $\left\{C^{(i)}, i = 1, 2, \ldots, n_C\right\}$ of $\mathbb{C}_d$. Thus a function, $L(\boldsymbol{q})$ proportional to the full likelihood, defined by

$$L(\boldsymbol{q}) = E\left[p(C \mid \boldsymbol{q})\right] \tag{4.5}$$

is a function on which inference about $\boldsymbol{q}$ can be based. Therefore, instead of an exhaustive evaluation, we recommend a Monte Carlo evaluation; that is, we generate a large number, $N$ say, of tables independently and uniformly on $\mathbb{C}_d$, $C_1, C_2, \ldots C_N$ and then approximate the summation in (4.5) by the Monte Carlo average

$$\widehat{L}(\boldsymbol{q}) = \frac{1}{N} \sum_{i=1}^{N} p(C_i \mid \boldsymbol{q})$$

Sampling uniformly on $\mathbb{C}_d$ is straightforward, due to the nature of the constraints; we only have to maintain column totals, and meet the constraints independently for each column. As noted above, there are $\binom{b_{+s}}{d_{+s}}$ possible configurations for column $s$, and therefore

$$\prod_{s=0}^{t-1} \binom{b_{+s}}{d_{+s}}$$

tables in total. To obtain a simulated column, we select $d_{+s}$ positions without replacement from the $b_{+s}$ available.

We have found that in moderately sized problems ($t = 20$, say) the Monte Carlo approximation often converges adequately with $N = 5000$, and this is sufficient to facilitate numerical maximization of the likelihood. In fact, $N = 100$ often gives a good approximation.

This Monte Carlo approach motivates our subsequent method of analysis, that is a Bayesian analysis based on an augmented likelihood. We shall see that a Markov chain constructed jointly on the parameter space of $\boldsymbol{q}$ (or a related set of parameters) and on $\mathbb{C}_d$ will facilitate full Bayesian inference.

## 5 Bayesian analysis

Inferences can be obtained within a Bayesian framework by formulating a missing-data representation. An MCMC scheme can then be used to impute the missing values and simultaneously make inference about the model parameters. In the present case, if only the totals, $d_{rs}$, are recorded, the missing values comprise the individual records, the rows of $C$. First, we utilize a reparameterization from the failure probabilities to a set of hazard probabilities, as this simplifies our Bayesian approach considerably.

### 5.1 Hazard parameterization and augmented likelihood

Bayesian inference is facilitated via a hazard parameterization. We now demonstrate how the likelihood can be simplified using the missing-data representation based on $C$, whose constituent matrices are the $C_{r+1}$ given above. The likelihood function for the augmented data, $C$, is relatively simple, as all the individual lifetimes and censoring times are known.

For the augmented data likelihood, suppose that there are, in total, $n_j$ lifetimes of length $j$ and $m_j$ lifetimes right-censored at $j$ ($j = 1, \ldots, t$) across all components ($i = 1, \ldots, b_r$) in all cohorts ($r = 0, \ldots, t-1$). Then the missing-data or augmented likelihood is

$$p(C \mid \boldsymbol{q}) = \prod_{j=1}^{t} q_j^{n_j} Q(j)^{m_j}$$

In terms of the hazard probabilities, this is

$$\prod_{j=1}^{t}\left\{h_j\prod_{k=1}^{j-1}(1-h_k)\right\}^{n_j}\left\{\prod_{k=1}^{j}(1-h_k)\right\}^{m_j}=\prod_{j=1}^{t}{h_j}^{n_j}(1-h_j)^{w_j}$$

where $w_t = m_t$ and

$$w_j=\sum_{k=j+1}^{t}(n_k+m_k)\quad \text{for}\quad j=2,\ldots,t-1.$$

In the hazard parameterization, therefore, the likelihood is effectively a product of independent binomial terms. This derivation is, of course, similar to that of the Kaplan-Meier estimator of survival. Inference will now proceed for $h_1, h_2, \ldots$, rather than for $q_1, q_2, \ldots$; subsequent reparameterization to $q_1, q_2, \ldots$ will then follow in straightforward fashion in the simulation-based approach that we use.

### 5.2 *Prior specification and truncation coherence*

We consider two forms of prior distribution, and then consider prior coherency issues by noting the use of anchored prior distributions.

#### 5.2.1 *The Dirichlet Prior*

A Dirichlet prior distribution for $\boldsymbol{q}$, $\boldsymbol{q} \sim \mathit{Dirichlet}(\alpha_1, \ldots, \alpha_{t+1})$, is given by

$$p(q_1,\ldots,q_t)=\frac{\Gamma\left(\alpha_1+\ldots+\alpha_{t+1}\right)}{\prod_{s=1}^{t+1}\Gamma(\alpha_s)}\left\{\prod_{s=1}^{t}q_s^{\alpha_s-1}\right\}Q(t)^{\alpha_{t+1}-1}. \tag{5.1}$$

This corresponds to a prior for $\boldsymbol{h}$ in which $h_1, \ldots, h_t$ are independent with marginal distributions

$$h_j \sim \mathit{Beta}\left(\alpha_j, \alpha_{j+1}+\ldots+\alpha_{t+1}\right)\qquad (j=1,\ldots,t).$$

The Dirichlet prior is commonly used for multinomial data, but is not conjugate to a likelihood based on the failure probabilities derived from censored data. However, it is conjugate for a likelihood in the hazard parameterization and this is the principal reason for using that parameterization. A more natural, and flexible, conjugate prior for the discrete hazards would again have $h_j$ *a-priori* independent, but with $h_j \sim \mathit{Beta}(\alpha_j, \beta_j)$ for $j = 1, \ldots, t$, as suggested by Mosiman (1962).

#### 5.2.2 *Truncation coherence*

The Dirichlet prior for $\boldsymbol{q}$ is attractive, due to its tractability, but some care needs to be taken in the choice of the hyperparameters $(\alpha_1, \ldots, \alpha_{t+1})$. We wish to express our prior opinion for the countably infinite set of probabilities $(q_1, q_2, \ldots)$, without prior knowledge of when truncation will occur, that is, a prior distribution on the space of discrete distributions on the positive integers. The truncation issue is important as we may wish to combine data from different sources where the underlying lifetime distribution is the same, but the truncation mechanism used differs from source to source. We also feel that a truncation coherent prior is more intuitively satisfying.

Prior distributions on probability and distribution functions have been much discussed in the Bayesian nonparametrics literature, and we use similar ideas to construct our prior here. Note, however, that we are restricting attention to priors for **discrete** distributions over a countable set of possible values, so many aspects of the nonparametric specification are simplified. We make the connection with our work though a discussion of the Dirichlet process (Ferguson (1973)) defined as follows. Let $(\Theta, B)$ be a measurable space, let $F_0$ be a probability measure on the space, and let $\alpha$ be a positive real number. A Dirichlet process, $DP(\alpha, F_0)$, is any distribution of a random probability measure $F$ over $(\Theta, B)$ such that, for all finite partitions $(A_0, \ldots, A_r)$ of $\Theta$,

$$(F(A_0),\ldots,F(A_r))\sim \mathit{Dirichlet}(\alpha F_0(A_0),\ldots,\alpha F_0(A_r)).$$

Parameter $\alpha$ is the precision parameter, and measure $F_0$ is the baseline or anchoring measure. This definition encompasses distributions on quite general measurable spaces. In this paper, the components of the finite partitions of interest are extremely straightforward: we have $A_j \equiv \{j\}$ for integer $j$, as we are dealing with discrete lifetime distributions only. Thus the anchoring measure $F_0$ is merely a discrete probability measure on the positive integers, with mass function $f_0$.

For $j = 1, 2, \ldots$, we let $\alpha_j = \alpha f_0(j) = \alpha\pi_j$, say, where $\alpha > 0$, and $\pi_1, \pi_2, \ldots$ is a given probability distribution on $\{1, 2, \ldots\}$. The advantage of this formulation is that we can use a familiar parametric form for the $\pi_j$: for example,

- geometric: $\pi_j = \theta^{j-1}(1-\theta)$
- logarithmic: $\pi_j = -\theta^j/\{j\log(1-\theta)\}$
- truncated-Poisson: $\pi_j = e^{-\lambda}\lambda^j/\{(1-e^{-\lambda})j!\}$

Alternatively, we can use a discretized version of a continuous model, such as the Weibull: $\pi_j = F(j) - F(j-1)$, where $F$ is the specified distribution function. For the Dirichlet distribution, with this choice of hyperparameters, we have

$$\mathrm{E}(q_j) = \pi_j \quad \text{and} \quad \mathrm{var}(q_j) = \frac{\pi_j(1-\pi_j)}{(\alpha+1)}. \tag{5.2}$$

The parameter $\alpha$ controls the variability in the prior; increasing $\alpha$ encourages the prior on $(q_1, q_2, \ldots)$ to be nearer the anchoring distribution, and if $\alpha$ is near 0 there is relatively large prior variability.

The advantage of using a prior structured in the way described above is that it is coherent across different levels of data-truncation. In the present application censoring occurs for the $r$th cohort of components at time $t-r$ $(r = 0, \ldots, t-1)$. On the other hand, a Dirichlet prior with $\alpha_j = 1$ for $j = 1, \ldots, t+1$, which might be selected as the default non-informative prior, is not consistent with a similar one adopted for a different value of $t$. For example, $\mathrm{E}(q_j \mid \text{follow-up to time } t) = 1/(t+1)$, which depends on $t$. The truncation coherent prior, when truncation is at time $t$, is Dirichlet with parameter vector

$$(\alpha_1, \alpha_2, \ldots, \alpha_t, \sum_{j=t+1}^{\infty} \alpha_j).$$

Also, in terms of the hazard parameterization, under this prior structure

$$\mathrm{E}(h_j) = \frac{\pi_j}{Q^{(\pi)}(j-1)} \quad \text{and} \quad \mathrm{var}(h_j) = \frac{\pi_j(1-\pi_j)}{Q^{(\pi)}(j-1)\{\alpha Q^{(\pi)}(j-1)+1\}}; \tag{5.3}$$

so, again, the hazard prior expectation and variance behave sensibly.

We have chosen not to elaborate the model further but adding another level to the hierarchy, although this is certainly a possible strategy. For example, rather than fixing $\boldsymbol{\pi}$ to be defined by the geometric discrete mass function with parameter $\theta$ fixed, we could place a hyperprior on $\theta$. Furthermore, the moment results in equations (5.2) and (5.3) allow us to consider particular functional forms for the expected probability or hazard, and then backsolve to obtain the appropriate prior specifications. For example, if we require *a priori*

$$\mathrm{E}(h_j) = \frac{\pi_j}{Q^{(\pi)}(j-1)} = g(j)$$

for some function $g$ on the non-negative integers (bounded by zero and 1), then we backsolve to obtain

$$\pi_j = g(j) \prod_{i=1}^{j-1} (1 - g(i))$$

and allow $\alpha$ to control the variability about this expectation. Finally, in particular circumstances, prior knowledge of the components under study can be used to specify the expected probabilities/hazards.

### 5.3 Posterior distribution

Assuming follow-up only until time $t$, with the truncation coherent prior described in the previous section, the joint posterior is given up to proportionality by

$$p(\boldsymbol{h} \mid C) \propto \prod_{j=1}^{t} h_j^{\alpha_j^*-1}(1-h_j)^{\beta_j^*-1},$$

where

$$\alpha_j^* = \alpha_j + n_j \quad \text{and} \quad \beta_j^* = \sum_{r=j+1}^{\infty} \alpha_r + w_j \qquad (j = 1, \ldots, t).$$

Hence, conditional on $C$, the $h_j$ are *a-posteriori* independent with

$$h_j | C \sim Beta(\alpha_j^*, \beta_j^*) \quad (j = 1, \ldots, t).$$

On integrating out $h_1, \ldots h_t$, the log of the normalizing constant in the joint posterior is

$$\sum_{j=1}^{t} \{\log\Gamma(\alpha_j^*) + \log\Gamma(\beta_j^*) - \log\Gamma(\alpha_j^* + \beta_j^*)\}.$$

Note that the calculation of the posterior quantities under the Mosiman prior is identical, with $\beta_j^* = \beta_j + w_j$ for each $j$.

## 6 Bayesian computation

The key to the Bayesian solution is to use Markov chain Monte Carlo (MCMC) to explore the joint space of probability vectors (the $t$-dimensional simplex) and life tables with fixed column totals, $\mathbb{S}^t \times \mathbb{C}_d$ say. The posterior distribution for $\boldsymbol{q}$ is formed via the augmented likelihood and the prior distribution for the unknown parameters. Let $\boldsymbol{d} = (d_{+1}, \ldots, d_{+t})$. Then, following the argument used in section 4.2, we have

$$p(\boldsymbol{q} \mid \boldsymbol{d}) \propto p(\boldsymbol{d} \mid \boldsymbol{q}) p(\boldsymbol{q}) = \left\{ \sum_{C \in \mathbb{C}_d} p(\boldsymbol{d}, C \mid \boldsymbol{q}) \right\} p(\boldsymbol{q}) = \left\{ \sum_{C \in \mathbb{C}_d} p(C \mid \boldsymbol{q}) \right\} p(\boldsymbol{q})$$

An alternative view of this posterior calculation has $C$ as a parameter in the model. Then the joint posterior of interest can be manipulated as follows:

$$p(\boldsymbol{q}, C \mid \boldsymbol{d}) = \begin{cases} 0 & C \notin \mathbb{C}_d \\ p(C \mid \boldsymbol{q}) p(\boldsymbol{q}) & C \in \mathbb{C}_d \end{cases}$$

Our strategy involves utilizing the hazard parameterization, then marginalizing the joint posterior out over the hazard probabilities, and constructing a Metropolis-Hastings algorithm on the space of missing tables, that is, we explore the marginal posterior distribution for $C$ on $\mathbb{C}_d$, denoted below by $\pi_d$. After the acceptance or rejection of a candidate table has been carried out, the new hazard/failure probabilities can be sampled directly from the conditional posterior distribution given in Section 5.3; for the hazard probabilities this distribution is the product of Beta distributions that may be sampled trivially. A related approach was used in Stephens and Crowder (2004) for the analysis of warranty data; in that case the cell centries in the upper triangle of a square, two-way table with fixed margins were imputed using MCMC with the objective of performing Bayesian inference for the warranty lifetime distribution.

We note here that the marginal **posterior** distribution for $C$ is, in general, **not** uniform. If it were, we could sample directly from it, and have no need for MCMC. The distribution is not uniform as we have proposed a proper lifetime distribution for the failure times.

### *6.1 MCMC moves*

For moves on the table space, $\mathbb{C}_d$, in the complete table there are $b_+$ rows and $t$ columns with column totals $d_{+s}$ ($s = 1, \ldots, t$) and some structural zeros. Any Metropolis-Hastings (MH) step should only propose moves that obey these constraints. An initial legitimate starting table must be constructed; this is straightforward, as we merely fill each column of the table independently in such a way that the column constraints are met. There are four obvious types of move.

I Entire table changes: this move involves independent sampling from the uniform distribution on $\mathbb{C}_d$. This is achieved, essentially, by using a random permutation within each of the columns of the augmented table (retaining the structural zeros). This global move may permit large changes in the likelihood function, but may only be accepted rarely.

II Column-changes: a column, $s$ say, is selected uniformly at random and filled with binary digits, independently, subject to the column-sum constraints. There are $\binom{b_{+s}}{d_{+s}}$ possible configurations for column $s$, and we sample uniformly from them. The proposed column is accepted/rejected according to the usual MH acceptance probability; in this case that reduces to the likelihood ratio.

III Element-changes of entries in a column: a column, $s$ say, is selected uniformly at random, and one of the $d_{+s}$ non-zero column entries is selected uniformly at random and a randomly chosen pair, a 1 and a 0, are swapped. Again, the MH acceptance probability reduces to the likelihood ratio. This move is a local (incremental) version of Type I.

IV Element-switches: a pair of columns, $s_1$ and $s_2$, say, is selected at random and one non-zero element in each column is chosen. Suppose that element $k_1$ of column $s_1$ and element $k_2$ of column $s_2$ are selected: if the corresponding entries in $C$ are $C_{k_1 s_1}$ and $C_{k_2 s_2}$, respectively, the proposed new table is chosen to be identical to the current one but with

$$\begin{array}{ll} C_{k_1 s_1} = 0 & C_{k_1 s_2} = 1 \\ C_{k_2 s_1} = 1 & C_{k_2 s_2} = 0 \end{array}.$$

This move preserves the column totals (we have added and subtracted 1 from the column totals for columns $s_1$ and $s_2$) and the other table constraints.

Move IV is the most "local" of the three types. Moves I and II require re-calculation of the likelihood for each of the components corresponding to the first $b_{+s}$ rows of $C$, whereas Move IV only involves a likelihood change for components $k_1$ and $k_2$.

In each case the MH step described above has acceptance probability for candidate table $C'$ and current table $C$ given by

$$\min\{1, \pi_d(C')/\pi_d(C)\}. \tag{6.1}$$

We justify this in the following discussion. Recall the general Metropolis-Hastings formulation: if the target distribution $\pi_d(C)$ is to be sampled, and the Metropolis-Hastings proposal density for moving from $C$ to $C'$ is denoted $P(C, C')$, then the acceptance probability is

$$\min\left\{1, \frac{\pi_d(C')P(C', C)}{\pi_d(C)P(C, C')}\right\}.$$

For Move I, $P(C, C') = P(C')$ and is equal to the prior distribution over the table space, which is uniform, and hence

$$\frac{P(C', C)}{P(C, C')} = \frac{P(C)}{P(C')} = 1.$$

For Move II, $P(C, C') = P_s(C'_s)P_{(s)}(C_{(s)})$ where $C_s$ relates to column $s$, and $C_{(s)}$ denotes the remaining columns of the table; note that $C'_{(s)} = C_{(s)}$, as the columns apart from $s$ are left unchanged. Therefore for Move II,

$$\frac{P(C', C)}{P(C, C')} = \frac{P_s(C_s)P_{(s)}(C_{(s)})}{P_s(C'_s)P_{(s)}(C_{(s)})} = \frac{P_s(C_s)}{P_s(C'_s)} = 1$$

as column $s$ is being sampled from its uniform prior distribution. For Move III, the proposal kernel is **symmetric**, that is

$$P(C, C') = P(C', C) \qquad \therefore \qquad \frac{P(C', C)}{P(C, C')} = 1.$$

Finally, for Move IV, the proposal kernel is again symmetric, and the same argument applies. Thus, in each case the Metropolis-Hastings ratio reduces to the ratio of marginal posterior quantities, as in equation 6.1. The only other check that we must carry out is that the Markov chain is irreducible on the table space, but the presence of Move I in the mixture kernel and the feasible starting value, ensures this, and hence ensures that the stationary distribution of the chain is the correct target distribution, $\pi_d$, by standard theory (see, for example, Tierney (1994)).

### 6.2 A Note on Censoring and Auxiliary variables

A referee has raised the issue of overcoming problems with censoring by including the unobserved failure times of components as further missing data, and including updates for these auxiliary parameters as part of the MCMC routine. This is certainly a possible approach to follow, and is similar to the one used in the analysis of, for example, competing risks data (see for example, Crowder (2001)). In this case, however, as there is potentially an auxiliary variable for each row in the table, such an approach does not seem so attractive, especially as the model is reasonably analytically tractable in the hazard parameterization.

### 6.3 A small example

Consider the following example in which $t = 5$, $b = 1$ for each $r$, and $d_+ = (d_{+1}, d_{+1}, d_{+3}, d_{+4}, d_{+5}) = (0, 1, 2, 2, 3)$. The initialized missing data table might be of the following form where the $*$ indicates the position of a structural

Table 3: Example Table

| $r$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 1 | 1 |
| 1 | * | 1 | 0 | 0 | 0 |
| 2 | * | * | 1 | 1 | 1 |
| 3 | * | * | * | 0 | 1 |
| 4 | * | * | * | * | 0 |
| Total | 0 | 1 | 2 | 2 | 3 |

zero. This table will have been obtained, for $s = 1, 2, 3, 4, 5$, by sampling $d_{+s}$ positions without replacement from the rows $1, \dots s$. For computing the likelihood, we store the summary statistics: these are the number, $n_j$, of observed lifetimes of length $j$, $j = 1, 2, \dots, 5$ indicated by the table, and the number, $m_j$, of censored lifetimes of each length. In this case, we have the summary table given in Table 4 and there are a total of

$$\binom{1}{0}\binom{2}{1}\binom{3}{2}\binom{4}{2}\binom{5}{3} = 1 \times 2 \times 3 \times 6 \times 10 = 360$$

tables that meet the column constraints. However, by exhaustive enumeration, it can be shown that there only 21

Table 4: Counts table for the small example

| $j$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $n_j$ | 6 | 1 | 1 | 0 | 0 |
| $m_j$ | 1 | 0 | 1 | 0 | 0 |

possible configurations of the summary statistics (one of which is given in Table 4). The likelihood, therefore, is a sum of the augmented likelihoods evaluated conditional on the 21 sets of summary statistics in turn, with a multiplier that corresponds to the number of tables that give rise to that set of summary statistics.

For an illustrative analysis, we use Markov chain Monte Carlo on the parameter space for $\boldsymbol{q}$ after the exact likelihood is computed by summing over the 21 tables. We utilize the hazard parameterization, and implement independent proposals for $(h_1, \ldots, h_5)$, that is, we propose moves for all five hazard parameters uniformly on the unit interval. For illustration, we retain a uniform Dirichlet prior on $\boldsymbol{q}$, corresponding to a prior on $\boldsymbol{h}$ of the form

$$p(\boldsymbol{h}) = p(h_1, \ldots, h_5) = (1-h_1)^4(1-h_2)^3(1-h_3)^2(1-h_4).$$

Moves in the independence Markov chain from $\boldsymbol{h}$ to $\boldsymbol{h}'$ are accepted with probability

$$\min\left\{1, \frac{l(\boldsymbol{h}' \mid \boldsymbol{d})p(\boldsymbol{h}')}{l(\boldsymbol{h} \mid \boldsymbol{d})p(\boldsymbol{h})}\right\}$$

as the proposal density is constant on $(0,1)^5$. In this analysis, the chain was run for two hundred thousand iterations, with every hundredth sample stored after a burn in of ten thousand iterations. The acceptance rate was approximately 43 % for the independence Metropolis chain. The results of this analysis are contained in Table 5, and Figures 1 and 2. It is clear from the histograms and scatter plots that there is some evidence of multimodality in the posterior distribution of the lifetime probabilities. In this instance, the independence Metropolis sampler in the hazard parameterization is capable of moving between the modes with relative ease. The general issue of posterior multimodality, and how it is addressed in a MCMC analysis, is discussed in the next section.

Table 5: Posterior summaries for the small example

| Quantile | $q_1$ | $q_2$ | $q_3$ | $q_4$ | $q_5$ |
|---|---|---|---|---|---|
| 2.5 % | 0.159 | 0.004 | 0.001 | 0.000 | 0.000 |
| 50 % | 0.570 | 0.152 | 0.057 | 0.029 | 0.012 |
| 97.5 % | 0.892 | 0.770 | 0.345 | 0.232 | 0.148 |

### *6.4 Exploring the modes of the posterior*

One common problem in using MCMC moves that are local in nature is that, if the posterior distribution is multimodal, it is possible that a single chain will get trapped in one region of high posterior probability and not explore the other high probability regions. We have found that in smaller examples, this appears not to be a serious problem. However, for larger examples, there is some cause from concern: chains run from different randomly-chosen starting tables attain significantly different log-posterior values after a reasonably large number of iterations.

This issue is relevant for the Baxter (1994) data, which has an unobserved table with $4145 \times (58 \times 57)/2 = 6851685$ cells. As an attempt to overcome this problem, we would adopt the following pragmatic strategy. We run $N_C$ chains from different starting tables and, after a given burn-in period, collect $M$ samples of the hazard probability vectors from each chain with suitable burn-in period, with the log-posterior probability for the unobserved table recorded for each sample. It will be evident if there is multimodal behaviour, as the log-posterior values can be compared directly. In any case, we can approximate the posterior distribution of the hazard probabilities using a Monte Carlo average over all chains, essentially assuming that those tables that lie in the union of the chains' sample paths comprise the entirety of the table sample space. Equivalently, we can re-sample $M$ of the $N_C \times M$ sampled probability vectors with resampling weights determined by the recorded log-posterior; in the discrete table space, the required normalization step is straightforward.

We have also explored other MCMC strategies that enable us to explore all corners of the table space $\mathbb{C}_d$. Two methods are of particular interest. The first is simulated annealing, where the Metropolis algorithm described above is used to locate the mode of the posterior distribution on the table space. Specifically, the algorithm is run on distribution $\pi_d^{1/T}$ for "temperature" $T$: as $T \to \infty$ over successive iterations, the distribution tends to become more peaked, so that eventually the Markov chain only explores the vicinity of the mode of $\pi_d$. The second method is an MCMC approach specifically designed to explore multimodal posteriors, and is based on the simulated annealing algorithm. The generation of candidate table $C'$ is achieved by a using a sequence of $N$ intermediate Metropolis steps and tables

$$C \to C^{(1)} \to C^{(2)} \ldots \to C^{(N)} \to C',$$

where the Metropolis step $C^{(r-1)} \to C^{(r)}$ is via the Metropolis kernel $\pi_d^{1/T_r}$. The temperature sequence $\{T_r\}$ is chosen such that the proposal distributions are gradually less peaked, and then more peaked, to allow moves between modes in a multimodal distribution.

The approach described above is known as *tempering* - see, for example, Neal (1996), and is just one of a number of techniques that can be used to explore awkward distributions. We have found that tempering, population and evolutionary MCMC (Liang and Wong, 2001) and sequential Monte Carlo (SMC) (see for example Doucet et al. 2001) are particularly useful; in other work, we have extensively studied the use of such algorithms in the analysis of mixture and clustering problems - see for example, Jasra et al. (2005), section 3.2. For general state spaces, population methods appear to be extremely promising; we can prove uniform ergodicity of population MCMC approaches (Jasra et al. 2005), and in the mixture context have shown that population MCMC methods often have superior performance compared to more straightforward samplers. In the context of this paper, where the state-space is finite, tempering appears to work well.

## 7 Prediction

A key objective in quasi life table analysis is prediction: if observational data are available up to time $t$, it is of keen interest to make predictive inferences about the numbers of failures (and hence repairs) that will be required at times $t+1, t+2, \ldots$. These will involve components currently in operation (that is, installed at times $0, \ldots, t-1$), plus components installed subsequently.

Recall that $b_r$ components are installed at time $r$ ($r = 0, \ldots, t$), and suppose that forecasts are required for the numbers of repairs or replacements that will be required in the future. We assume that $b_{t+1}, \ldots, b_{t+s}$ are known or can be projected. Then we wish to make predictive inferences about the column totals in an extended table with $t+s$ columns.

Table 6: Data observed and to be predicted

| Time | 1 | 2 | ... | $t$ | $t+1$ | $t+2$ | ... | $t+s-1$ | $t+s$ |
|---|---|---|---|---|---|---|---|---|---|
| $b_0$ | $d_{01}$ | $d_{02}$ | ... | $d_{0t}$ | $d_{0,t+1}$ | $d_{0,t+2}$ | ... | $d_{0,t+s-1}$ | $d_{0,t+s}$ |
| $b_1$ | * | $d_{12}$ | ... | $d_{1t}$ | $d_{1,t+1}$ | $d_{1,t+2}$ | ... | $d_{1,t+s-1}$ | $d_{1,t+s}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $b_{t-1}$ | * | * | ... | $d_{t-1,t}$ | $d_{t-1,t+1}$ | $d_{t-1,t+2}$ | ... | $d_{t-1,t+s-1}$ | $d_{t-1,t+s}$ |
| $b_t$ | * | * | ... | * | $d_{t,t+1}$ | $d_{t,t+2}$ | ... | $d_{t,t+s-1}$ | $d_{t,t+s}$ |
| $b_{t+1}$ | * | * | ... | * | * | $d_{t+1,t+2}$ | ... | $d_{t+1,t+s-1}$ | $d_{t+1,t+s}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | ... | $\vdots$ | $\vdots$ | $\vdots$ | ... | $\vdots$ | $\vdots$ |
| $b_{t+s-2}$ | * | * | ... | * | * | * | ... | $d_{t+s-2,t+s-1}$ | $d_{t+s-2,t+s}$ |
| $b_{t+s-1}$ | * | * | ... | * | * | * | ... | * | $d_{t+s-1,t+s}$ |
| Total | $d_{+1}$ | $d_{+2}$ | ... | $d_{+t}$ | $d_{+,t+1}$ | $d_{+,t+2}$ | ... | $d_{+,t+s-1}$ | $d_{+,t+s}$ |

Table 6 illustrates the prediction task: row $r$ ($r = 0, 1, \ldots$) the total numbers of repairs at times $r+s$ ($s = 1, 2, \ldots$) for the $b_r$ components installed at time $r$. The entries in the right-hand half of the table, beyond time $t$, are to be predicted.

Prediction (i.e. guessing the value of as-yet-unobserved random variables) can be a computationally-complicated task. Let $D^*(t,s)$ denote the repairs-totals vector $(d_{+,t+1}, \ldots, d_{+,t+s})$ for years $(t+1, \ldots, t+s)$. Then, formally, the posterior predictive distribution is

$$p\{D^*(t,s) \mid d_{+1}, \ldots, d_{+t}\} = \sum_C \int p\{D^*(t,s) \mid C, \boldsymbol{q}\}\; p(C, \boldsymbol{q} \mid d_{+1}, \ldots, d_{+t})\; d\boldsymbol{q},$$

where $p(C, \boldsymbol{q} \mid d_{+1}, \ldots, d_{+t})$ is the posterior distribution. This expression is not generally computationally feasible. In an MCMC setting, however, prediction is more straightforward. We examine the full conditional

$$p\{D^*(t,s) \mid d_{+1}, \ldots, d_{+t}, C, \boldsymbol{q}\},$$

which can be decomposed by examining the different rows of $C$. Thus, the predicted failure totals can be sampled by sampling and summing over the individual failure profiles for each component in all of the cohorts from time 0 to time $t+s-1$ inclusive.

For an item in cohort $r$ ($r = 0, \ldots, t-1$) the predicted failure profile must be sampled conditionally on $C$; for component $i$ in the $r$th cohort, conditional on $C$, it is known that the $(m_i^{(r)}+1)$st lifetime is greater than $l^{(r)}_{i,m_i^{(r)}+1} = k$, say, as in the construction in Section 5. Thus, the next failure time will occur according to the failure distribution defined by $(q_1, q_2, \ldots)$ restricted to the range $(k+1, k+2, \ldots)$, that is

$$\frac{q_{k+1}}{Q(k)}, \frac{q_{k+2}}{Q(k)}, \ldots$$

This distribution may be sampled using inversion of the distribution function. For all subsequent lifetimes for this component, the failure distribution $(q_1, q_2, \ldots)$ should be used, which is also the case for items in cohorts $r = t, \ldots, t+s-1$.

For prediction from the full conditional distribution, we require the current values of $q_s$ for all $s = 1, 2, \ldots$, and in the posterior analysis above we have only sampled $q_1, \ldots, q_t$. Sampling the values for $q_{t+1}, q_{t+2}, \ldots$, conditional on $q_1, q_2, \ldots, q_t$, is straightforward: we sample recursively, for $s = t+1, t+2, \ldots$

$$q_s \mid q_1, \ldots, q_{s-1}, C \sim Beta\left(\alpha_s, \sum_{j=s+1}^{\infty} \alpha_j\right) \qquad \text{restricted to } (0, Q(s-1))$$

that is, essentially, the successive conditional priors on $q_{t+1}, q_{t+2}, \ldots$. For predictions up to time $t+s$ we need only sample values up to $q_{t+s}$.

The capability of our model to encompass prediction in a natural fashion as part of the MCMC procedure is an attractive facet. Other methods of prediction, for example the chain-ladder method (see, for example Verrall (2000), Mack and Venter (2000)) may also be implemented. See the further discussion of chain-ladder methods in contrast with our approaches in Stephens et al. (2004).

## 8 Examples and results

We give four examples: the first three are simulated, based on geometric and non-parametric models respectively; the final one is a discussion of the Baxter (1994) data given in Table 1.

### *8.1 Example 1: geometric data*

In our first example we study both the analysis of a small simulated data set, and then the empirical (frequentist) coverage performance of the posterior distribution (using the posterior credible interval) over a number of replicated data sets. We contrast the Bayesian results with those obtained with the asymptotic estimating-equations approach of Crowder and Stephens (2003).

First, a simulated data set was generated, for which $t = 20$, $b_r = 1$ ($r = 0, 1, \ldots, 19$) and the column totals $d_{+s}$ ($s = 1, \ldots, 20$) are

$$0, 1, 1, 1, 3, 5, 2, 3, 2, 4, 4, 4, 5, 5, 4, 6, 7, 6, 8, 7$$

The simulated data form a square, upper-trangular array, with column totals fixed; the number of failures of a particular component, and hence the row totals, are unspecified. These data were generated from a $Geometric(\theta)$ model where, for $j = 1, \ldots, t$,

$$q_j = \text{pr}(\text{Lifetime} = j) = (1-\theta)^{j-1}\theta \quad \text{and} \quad \text{pr}(\text{Lifetime} > j) = (1-\theta)^j = Q(j)$$

with $\theta = 0.4$, and represent a case where a single unit is commissioned each month (say), and the unit us liable to fairly frequent breakdown.

In the Geometric model we are able to make inference on the parameter $\theta$ directly. The matrix $C$ contains precisely $d_{tot} = \sum_{s=1}^{t} d_{+s}$ ones, and the remaining $\sum_{r=0}^{19}(t-r)b_r - d_{tot}$ entries are zero. In the Geometric model, the lack-of-memory property implies that the likelihood for $\theta$ is binomial, that is

$$D_{tot} \sim Binomial\left(\sum_{r=0}^{19}(t-r)b_r, \theta\right),$$

and hence the maximum likelihood estimate and posterior distribution for $\theta$ are available with knowledge of $d_{tot}$ only. Here, $d_{tot} = 78$ and $\sum_{r=0}^{19}(t-r)b_r = 20 \times 19/2 = 190$ so the maximum likelihood estimate is $\hat{\theta} = 78/190 = 0.411$ with estimated standard error $\sqrt{\hat{\theta}(1-\hat{\theta})/190} = 0.036$.

We now present the results of a Bayesian analysis of the data assuming the missing data likelihood outlined in previous sections using a non-parametric model. We wish to estimate the hazard probabilities $(h_1, \ldots, h_{20})$ or the failure probabilities $(q_1, \ldots, q_{20})$. For illustration in this example, a uniform prior distribution over the simplex is used, that is, we assume that the $\alpha$-parameters in the Dirichlet prior (5.1) each take the value 1. The results given below were generated from a single run with burn-in 50000 iterations, 5000 samples collected over the next 2.5 million iterations (a rather conservative sampling strategy) with total computation time less than one hour on a 1Gb RAM, 2.4GHz PC. Convergence was assessed by inspection of trace plots and autocorrelation functions; repeated runs gave virtually identical answers. Numerical posterior summaries are given in Table 7 for the first six failure probabilities. Figure 1 presents the results in the form of a boxplot of the sampled failure probabilities. It is clear that the MCMC approach is reconstructing the unknown failure distribution accurately in light of the data. We note here that, for these data, the estimating-equation approach used by Crowder and Stephens (2003) does not produce valid parameter estimates; this is understandable, given the asymptotic justification for that approach.

Table 7: Posterior numerical summaries for Example 1

| Parameter | $q_1$ | $q_2$ | $q_3$ | $q_4$ | $q_5$ | $q_6$ |
|---|---|---|---|---|---|---|
| True value | 0.400 | 0.240 | 0.144 | 0.086 | 0.052 | 0.031 |
| mle | 0.453 | 0.248 | 0.136 | 0.074 | 0.041 | 0.022 |
| 2.5% quantile | 0.283 | 0.027 | 0.005 | 0.013 | 0.002 | 0.001 |
| 25% quantile | 0.373 | 0.084 | 0.038 | 0.055 | 0.016 | 0.011 |
| 50% quantile | 0.433 | 0.128 | 0.071 | 0.083 | 0.032 | 0.023 |
| 75% quantile | 0.488 | 0.178 | 0.108 | 0.118 | 0.053 | 0.041 |
| 97.5% quantile | 0.572 | 0.265 | 0.186 | 0.195 | 0.105 | 0.091 |

We now study coverage performance of the Bayesian procedure. Table 8 contains summaries of the 95 % coverage intervals for the Bayesian posterior median estimator for the first six probabilities in the lifetime distribution, derived from 2000 replications of the MCMC procedure on independent simulations from the Geometric model.

Table 8: Posterior coverage of Bayesian median estimator for Example 1

| Parameter | $q_1$ | $q_2$ | $q_3$ | $q_4$ | $q_5$ | $q_6$ |
|---|---|---|---|---|---|---|
| True value | 0.400 | 0.240 | 0.144 | 0.086 | 0.052 | 0.031 |
| Lower | 0.279 | 0.075 | 0.018 | 0.002 | 0.000 | 0.001 |
| Median | 0.417 | 0.152 | 0.101 | 0.024 | 0.012 | 0.018 |
| Upper | 0.522 | 0.294 | 0.206 | 0.123 | 0.096 | 0.070 |

This table demonstrates that the frequentist performance of the Bayesian procedure is good adequate for this relatively small data set.

### 8.2 Example 2: Geometric data and prior sensitivity

The second example is a simulated data set for which $t = 20$, $b_r = 20$ $(r = 0, \dots, 19)$ and the column totals $d_{+s}$ $(s = 1, \dots, 20)$ are

$$6, 8, 7, 19, 22, 18, 38, 28, 41, 36, 44, 55, 45, 61, 55, 69, 68, 80, 72, 85$$

giving $d_{tot} = 857$. These data were generated from a *Geometric*(0.2) lifetime model, and we have the mle $\hat{\theta} = 0.2255$ with estimated standard error 0.007.

The purpose of this example is to examine the sensitivity of inference to prior specification. Recall that we are aiming to make inference about the lifetime distribution (that is, twenty probabilities) from, essentially, twenty data (the column totals), and therefore might expect the prior to be relatively influential. We have studied the posterior for four different priors: the uniform prior and three more informative anchored priors:

1. : Uniform, Dirichlet (not truncation coherent).
2. : Anchored prior: anchoring to *Geometric*(0.20) with $\alpha = 20$.
3. : Anchored prior: anchoring to *Geometric*(0.20) with $\alpha = 200$.
4. : Anchored prior: anchoring to *Poisson*(10) with $\alpha = 20$.

Prior 3 is tightly anchored to the *Geometric*(0.2) distribution, and Prior 4 has a much longer expected lifetime. We present the results of a single run using the sampling strategy as above, and present the results in graphical and numerical form in Table 9 and Figure 2.

The boxplots in Figure 4 demonstrate the influence of the prior, but also that the likelihood contribution is not overwhelmed. For example, for prior 4, which is a fairly strongly anchored prior on a distribution that is in conflict with the true model, the correct geometric distributional shape is recovered for the first few lifetime probabilities.

### 8.3 Example 3: Non-Geometric Lifetime distribution

The third example is a simulated data set for which $t = 10$, $b_r = 200$ $(r = 0, \dots, 9)$ and the lifetime distribution has the form

$$q = (0.01, 0.04, 0.1, 0.1, 0.2, 0.3, 0.125, 0.05, 0.03, 0.02)$$

with remainder probability 0.025. The column totals $d_{+s}$ $(s = 1, \dots, 10)$ are

$$2, 11, 25, 61, 87, 150, 195, 230, 255, 296.$$

For this example, a truncation coherent, loosely-anchored geometric prior with $\alpha = 10$ and $\theta = 0.1$ was used.

Table 9: Posterior numerical summaries for Example 2

| | Parameter | $q_1$ | $q_2$ | $q_3$ | $q_4$ | $q_5$ | $q_6$ | $q_7$ | $q_8$ | $q_9$ | $q_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | True value | 0.200 | 0.160 | 0.128 | 0.102 | 0.082 | 0.066 | 0.052 | 0.042 | 0.034 | 0.027 |
| | MLE | 0.226 | 0.175 | 0.135 | 0.105 | 0.081 | 0.063 | 0.049 | 0.038 | 0.029 | 0.023 |
| | 2.5 % | 0.143 | 0.134 | 0.048 | 0.070 | 0.041 | 0.024 | 0.026 | 0.019 | 0.013 | 0.001 |
| Prior 1 | Median | 0.187 | 0.181 | 0.106 | 0.098 | 0.073 | 0.056 | 0.062 | 0.044 | 0.038 | 0.017 |
| | 97.5 % | 0.241 | 0.286 | 0.154 | 0.149 | 0.107 | 0.086 | 0.128 | 0.072 | 0.070 | 0.045 |
| | 2.5 % | 0.148 | 0.113 | 0.099 | 0.057 | 0.051 | 0.018 | 0.026 | 0.007 | 0.002 | 0.000 |
| Prior 2 | Median | 0.194 | 0.162 | 0.140 | 0.100 | 0.086 | 0.055 | 0.056 | 0.041 | 0.028 | 0.020 |
| | 97.5 % | 0.236 | 0.204 | 0.205 | 0.148 | 0.145 | 0.095 | 0.118 | 0.077 | 0.060 | 0.058 |
| | 2.5 % | 0.165 | 0.128 | 0.094 | 0.073 | 0.055 | 0.041 | 0.032 | 0.021 | 0.016 | 0.011 |
| Prior 3 | Median | 0.202 | 0.163 | 0.129 | 0.103 | 0.081 | 0.064 | 0.052 | 0.040 | 0.032 | 0.025 |
| | 97.5 % | 0.236 | 0.203 | 0.166 | 0.140 | 0.114 | 0.093 | 0.081 | 0.064 | 0.054 | 0.047 |
| | 2.5 % | 0.153 | 0.084 | 0.050 | 0.016 | 0.029 | 0.035 | 0.039 | 0.038 | 0.036 | 0.032 |
| Prior 4 | Median | 0.203 | 0.135 | 0.087 | 0.070 | 0.063 | 0.056 | 0.057 | 0.057 | 0.056 | 0.050 |
| | 97.5 % | 0.277 | 0.171 | 0.128 | 0.107 | 0.090 | 0.079 | 0.081 | 0.083 | 0.085 | 0.076 |

The results for this analysis are displayed in Figure 5 in the form of a boxplot for each of the ten lifetime probabilities. This plot (verified over replicate runs for the same data but different MCMC starting values) demonstrates that the posterior distribution is influenced by the prior, and does not recover the generating distribution. Under a uniform Dirichlet prior (results not shown here) the posterior distribution better reflects the generating model, but this example illustrates the level of difficulty of inference in this problem.

### *8.4 Example 4: Baxter data*

We have examined a subset of the data in Baxter (1994, Table 2), taking only the first two years-worth of data, so that $t = 58$. For each year, $b_r = 4145$ ($r = 0, \ldots, 57$) and the column totals are given here in Table 1. We carried out an MCMC analysis on a subset of these data, namely the first 24 of the 58 months.

We uncovered a previously hidden structure. Using standard Metropolis moves on the table space, and parallel MCMC runs, it became clear that the posterior distribution $\pi_d$ is multimodal. Runs starting from different initial configurations converged to one of a small number of modes. This impression of the posterior was confirmed using simulated annealing and tempered Metropolis as described above in section 6.4. The mode in which the highest posterior probability on $\mathbb{C}_d$ was obtained was revealing; it corresponds to a lifetime distribution that implies periodic replacement of the items after 12 months of operation, in that the estimated hazard probability for month 13 was virtually equal to 1. Given the context of the Baxter experiment, this result is entirely plausible. A complete posterior summary for this large table space is not possible here, but we note that the modes of the multimodal posterior distribution on the table space yield different lifetime distribution estimates. We also note that this "annual replacement" structure (plausible in hindsight) may not have been found by methods other than MCMC exploration.

## 9 Discussion

We have given a Bayesian formulation for the analysis of quasi life tables. Central to our approach has been an MCMC algorithm that samples the parameters of the unknown discrete lifetime distribution; our approach uses a data-augmentation approach as part of the MCMC to sample the unobserved lifetimes for the components conditional on the observed column totals.

In our analysis it is also recognized that, in the Bayesian framework, and in this inference problem in particular, the prior specification for the probabilities in the lifetime distribution plays an important role. We also pointed out that a standard prior specification based on the Dirichlet distribution is not coherent across different truncation times, and this motivated us to construct coherent priors based on ideas from Bayesian nonparametrics.

The data from different cohorts have been pooled on the assumption that the $q_l$ take the same values for different cohorts. If this is not the case, and the $q_l$ from different cohorts are unrelated, separate analyses will need to be performed. Otherwise, it may be that there is some specified relationship between the $q_l$ for different $r$, e.g. reflecting increasing component reliability over time, and then some appropriate pooling of data will be appropriate.

**Acknowledgements** We would like to thank the two referees and the coordinating editor for their helpful comments that have led to significant improvements in the presentation of the paper. The first author would like to thank the Department of Statistics, University of Washington, Seattle, for accommodating him whilst the revised version of the paper was being produced.

**References**

Baxter, L. A., "Estimation from quasi-life tables", *Biometrika* Vol. **81**, pp 567-577, 1994.

Crowder, M. J. *Classical Competing Risks*. Chapman and Hall/CRC, Boca Raton, FL, 2001.

Crowder, M. J. and Stephens, D.A., "On the analysis of quasi-life tables", *Lifetime Data Analysis* Vol. **9**, pp 345-355, 2003.

Doucet, A, de Freitas, N. and Gordon, N., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

Ferguson, T. S., "A Bayesian analysis of some nonparametric problems", *Annals of Statistics*, **1**, 209230, 1973.

Jasra, A, Holmes, C. C. and Stephens D. A., "Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling", *Statistical Science*, **20**, Pages: 50 - 67, 2005.

Jasra, A, Stephens, D. A. and Holmes, C. C., "Population-based reversible jump Markov chain Monte Carlo", Technical Report, Imperial College: available for download from `http://stats.ma.ic.ac.uk/das01/public_html/Papers.html`, 2005

Karim, M. R., Yamamoto, W. and Suzuki, K., "Statistical analysis of marginal count failure data", *Lifetime Data Analysis* **7**, pp173-186, 2001

Liang F. M., and Wong W. H. "Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models", *Journal of the American Statistical Association*, **96**: pp653-666.

Mack T., and Venter G., "A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates", *Insurance: Mathematics and Economics*, **26**(1), pp 109-111, 2000.

Mosiman, J. E., "On the compound multinomial distribution, the multivariate $\beta$-distribution and correlations among proportions", *Biometrika* Vol. **49**, pp 65-82, 1962.

Neal, R., "Sampling from multimodal distributions using tempered transitions", *Statistics and Computing*, Vol. **4**, pp 353-366, 1996.

Stephens D. A., and Crowder M. J., "Bayesian analysis of discrete time warranty data", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **53**, pp. 195-217, 2004.

Stephens D. A., Crowder M. J., Dellaportas P, "Quantification of automobile insurance liability: a Bayesian failure time approach", *Insurance Mathematics and Economics*, **34**, pp. 1-21, 2004.

Tierney, L. "Markov chains for exploring posterior distributions (with discussion)" . *The Annals of Statistics* **22**, pp. 1701–1762, 1994.

Verrall, R. J., "An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique". *Insurance: Mathematics and Economics*. **26**(1), pp 91-99, 2000.

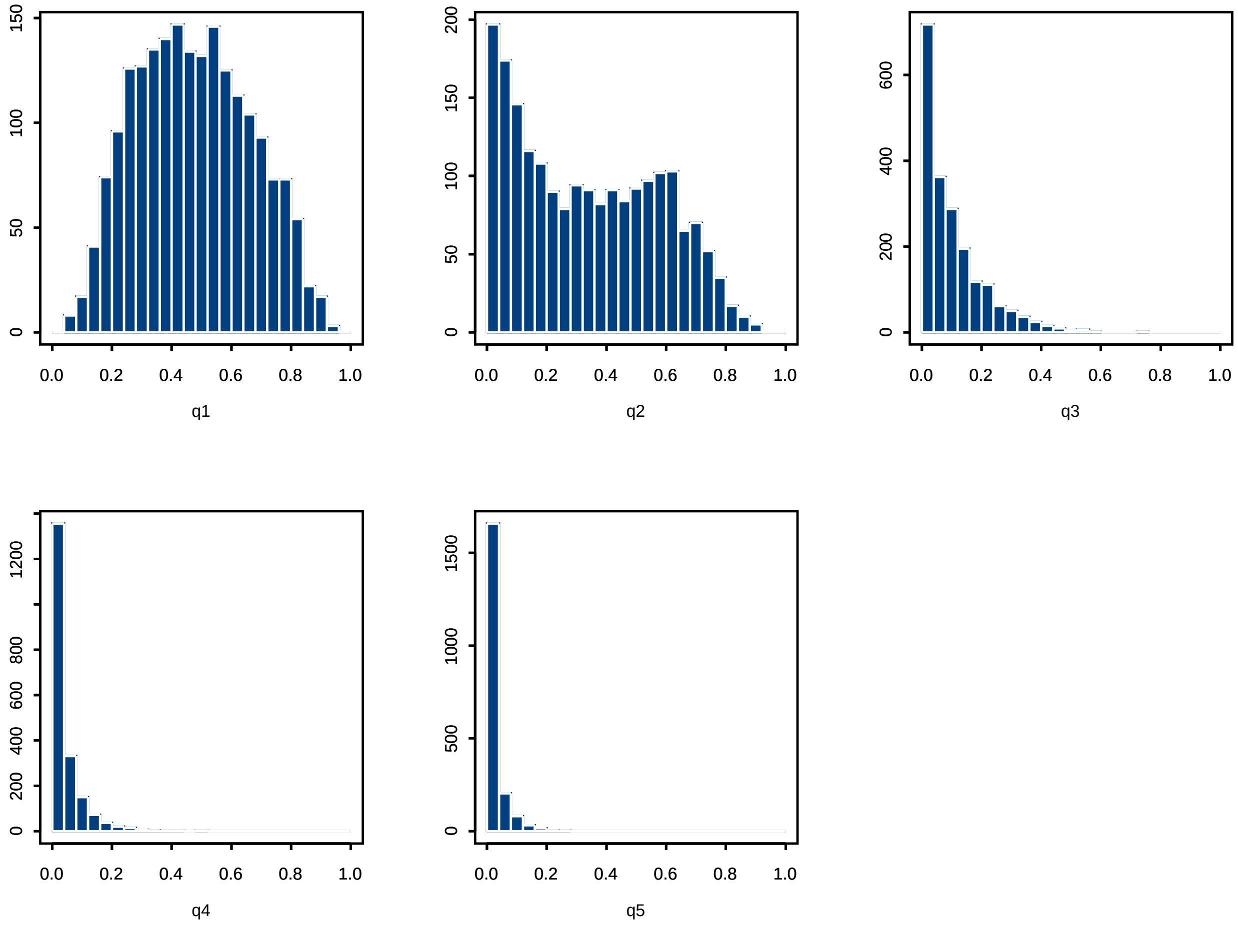


Figure 1: Analysis of the small example: Posterior histograms of lifetime distribution probabilities derived from MCMC analysis

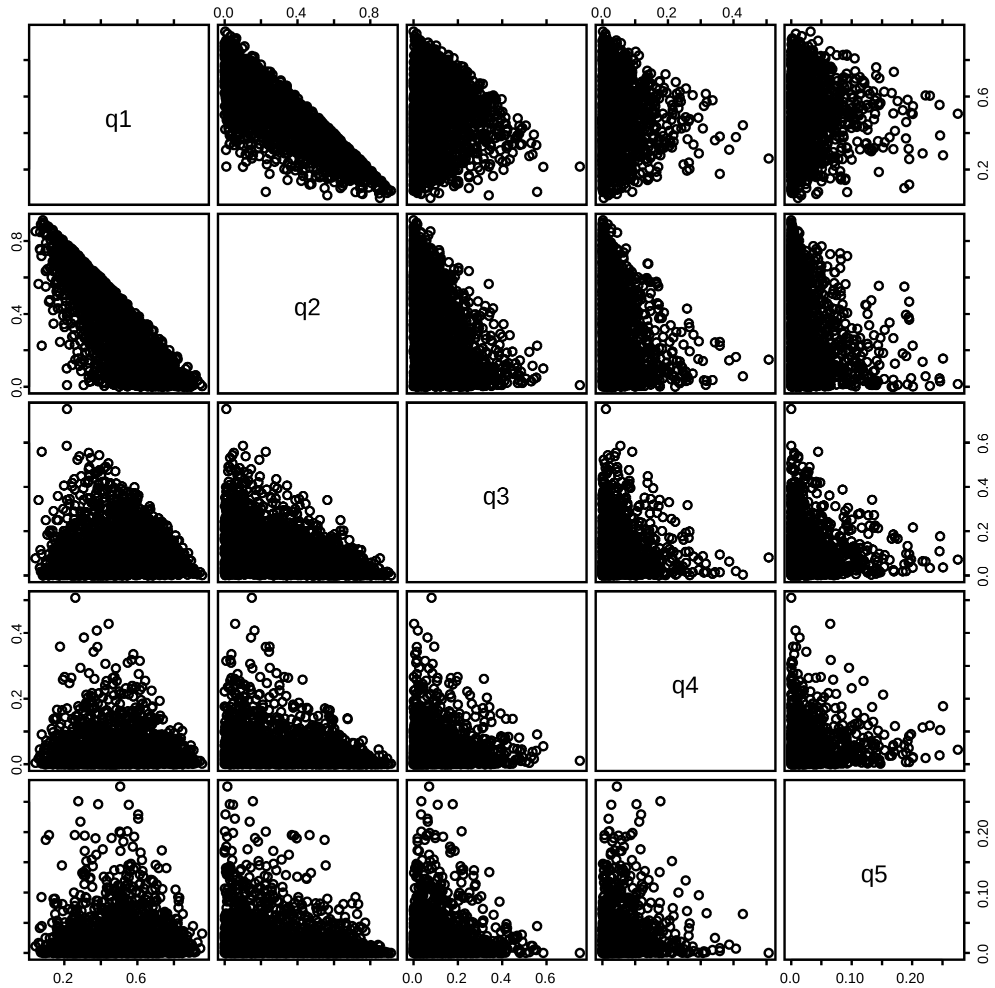


Figure 2: Analysis of the small example: Posterior scatterplot matrix

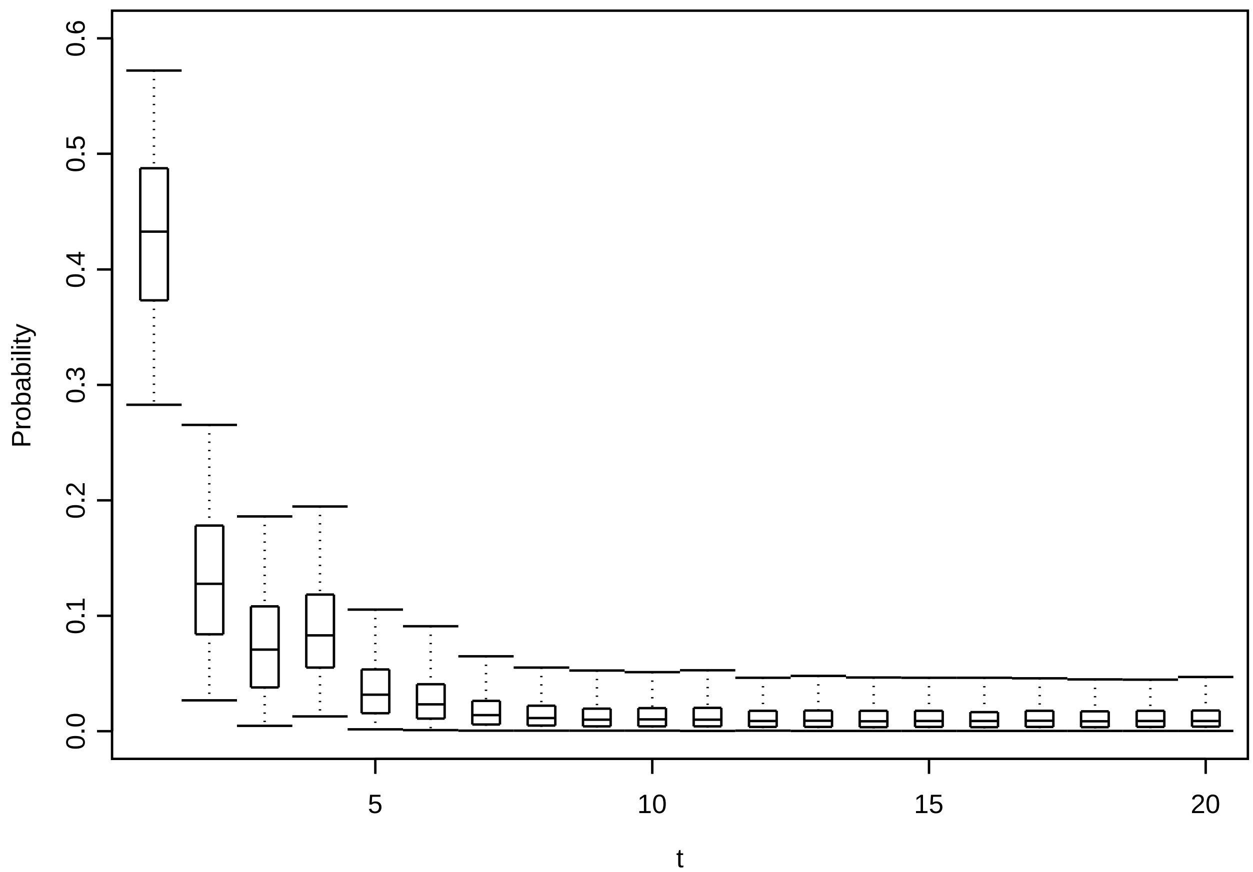


Figure 3: Analysis of Example 1: Posterior boxplots of lifetime distribution probabilities derived from MCMC analysis in the non-parametric model

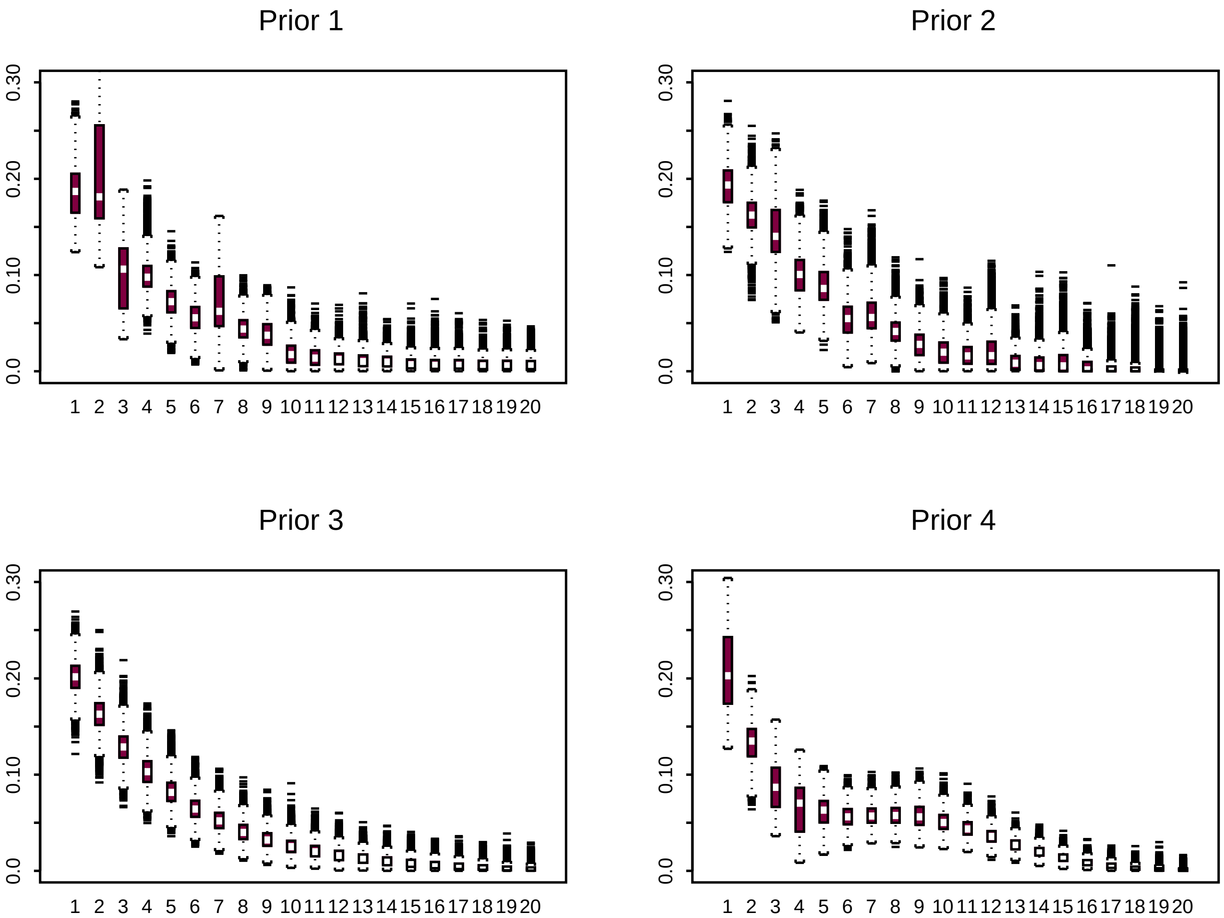


Figure 4: Analysis of Example 2: Posterior boxplots of lifetime distribution probabilities under the four priors for Example 2

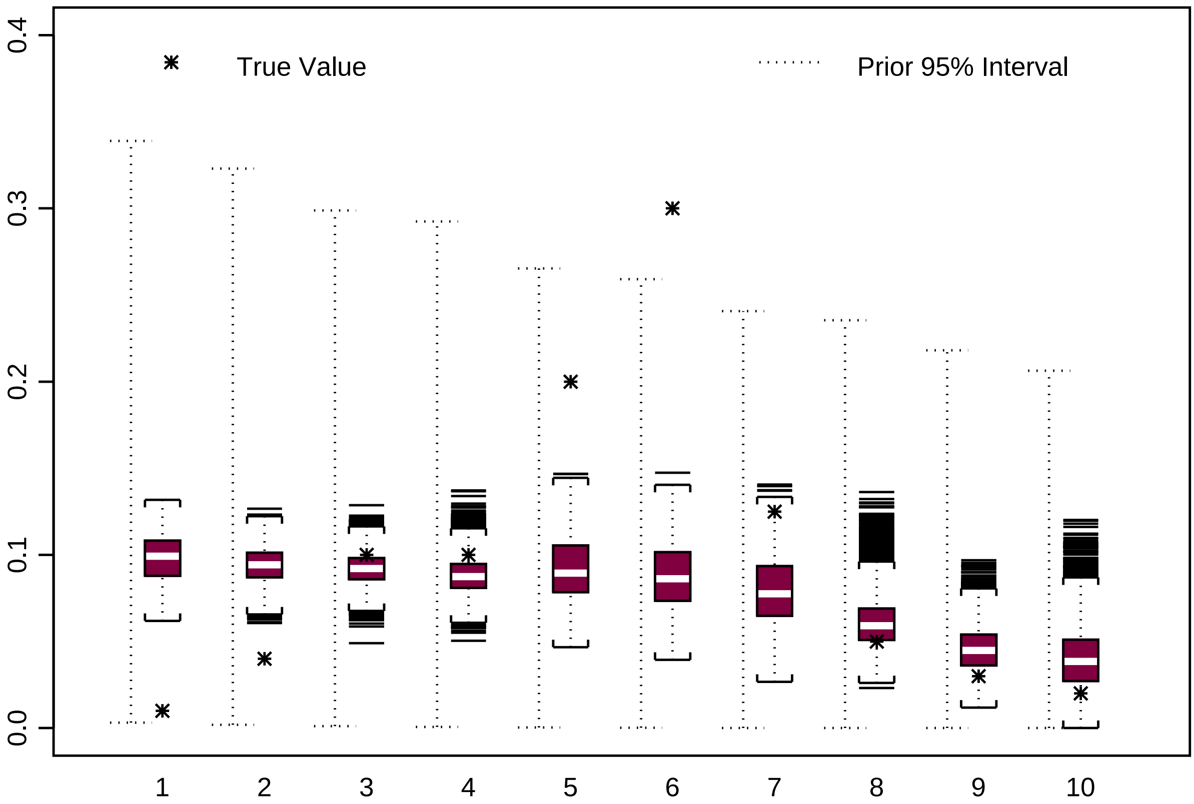


Figure 5: Analysis of Example 3: Posterior boxplots of lifetime distribution probabilities, with the true values and the 95 % Prior interval