Bayesian planning and inference of a progressively censored sample from linear hazard rate distribution

Ananda Sen

Departments of Family Medicine & Biostatistics

University of Michigan, Ann Arbor, Michigan, USA, anandas@umich.edu

Nandini Kannan

Department of Management Science and Statistics

University of Texas, San Antonio, Texas, USA, Nandini.Kannan@utsa.edu

Debasis Kundu

Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur, India, kundu@iitk.ac.in

Abstract

This paper deals with the Bayesian inference of the linear hazard rate (LHR) distribution under a progressively censoring scheme. A unified treatment of both Type I and Type II censoring is presented under independent gamma priors for the parameters, that yields the posteriors as mixtures of gamma. The priors are motivated from a *probability matching* viewpoint. Along with marginal inference and prediction, a joint credible set is constructed utilizing the posterior distribution of certain quantities of interest. The Bayesian inference demonstrates an intimate connection with the frequentist inference results under a Type-II censoring scheme. Bayesian planning strategies are explored that search for the optimal progressive censoring schemes under a variance criterion as well as a criterion based on the length of a credible interval for percentiles.

Key Words and Phrases: Linear hazard rate, Progressive censoring scheme; Bayesian inference; Markov Chain Monte Carlo; Joint credible set; Bayesian planning.

1 Introduction

In life-testing and reliability studies, a linear hazard rate (LHR) distribution is deemed useful in modeling the life-length of a system or component when failures occur at random and also from aging or wear-out. Structurally, it is the distribution of the minimum of two independent random variables, one having an exponential and the other having a Rayleigh distribution. The probability density function (pdf), cumulative distribution function (CDF) and hazard function (HF) of LHR(λ_1, λ_2), parameterized by $(\lambda_1, \lambda_2), \quad \lambda_1 > 0, \lambda_2 > 0$, are given by

$$f(t; \lambda_1, \lambda_2) = (\lambda_1 + 2\lambda_2 t)e^{-(\lambda_1 t + \lambda_2 t^2)}; \quad t > 0, \tag{1}$$

$$F(t; \lambda_1, \lambda_2) = 1 - e^{-(\lambda_1 t + \lambda_2 t^2)}, \quad t > 0, \tag{2}$$

and

$$h(t; \lambda_1, \lambda_2) = \lambda_1 + 2\lambda_2 t, \quad t > 0, \tag{3}$$

respectively. Exponential and Rayleigh distributions can be obtained as limiting cases of the LHR distribution by taking $\lambda_2 \longrightarrow 0$ and $\lambda_1 \longrightarrow 0$, respectively. A noteworthy fact is that this two parameter model in the increasing failure rate (IFR) class constitutes a generalization of the exponential distribution in the direction distinct from the gamma and Weibull. While, in the IFR case, both gamma and Weibull force the hazard rates to be zero at $t = 0$, LHR provides for $h(0) = \lambda_1 > 0$, a smoother and often more realistic transition from the constant failure rate to the IFR property. Another quite interesting feature of LHR is the fact that the residual life distribution of LHR follows yet another LHR. Specifically, if $T \sim LHR(\lambda_1, \lambda_2)$, then for any fixed $t_1 > 0$,

$$T - t_1 \mid T > t_1 \sim LHR(\lambda_1^*, \lambda_2), \quad \text{with} \quad \lambda_1^* = \lambda_1 + 2\lambda_2 t_1.$$

This is an extension of the lack-of memory property of Exponential but incorporating the ageing effect.

Potential of LHR as a life-model in reliability and survival applications had been demonstrated well by Broadbent (1958), Carbone et. al. (1967), and Kodlin (1967), (see also Chapter 4 of Gross and Clark, 1975). LHR can also be viewed as an approximation to other life-models such as Gompertz-Makeham distribution with hazard rate $h^*(t) = \gamma_0 + \gamma_1 e^{\gamma_2 t}$. Note that for small values of $\gamma_2 t$,

$$h^*(t) \approx \gamma_0 + \gamma_1(1 + \gamma_2 t) = \gamma_0' + \gamma_1' t,$$

matching the linear hazard form. Thus, a LHR will be an appropriate model in this case as long as one abstains from "long-term" predictions. Bain (1974) and Sen and Bhattacharya (1995) considered maximum likelihood estimation of LHR parameters for Type-II censored sample and Lin *et al.* (2003) discussed the estimation of the unknown parameters based on record and inter-record times. Balakrishnan and Malik (1986) established the recurrence relations for moments of order statistics from the LHR distribution. Various distributional properties and different applications of the LHR distribution have been summarized in the review article Sen (2006). Ashour and Youssef (1991) first considered the Bayes estimation of the unknown parameters of the LHR distribution. Unfortunately their derivation of the marginal posterior distributions is not correct (see Lin *et al.* 2006). Pandey *et al.* (1993) and recently Lin *et al.* (2006) considered Bayesian estimation in a more general set up.

The main objective of this article is to carry out Bayesian inference and planning for a LHR distribution under a progressive censoring (Type-I or Type-II) scheme of sampling. In life-testing experiments, one often encounters situations where it takes a substantial amount of time to obtain a reasonable number of failures necessary to carry out reliable inference. In such situations, traditional Type-I or Type-II censoring schemes would tie up the units on test, which could potentially be used for other types of experiments. Motivated by the need of using the same units in multiple testing running in parallel, a censoring strategy called

Progressive Censoring (PC), introduced by Cohen (1963), has found considerable attention in the past four decades. PC is a method that enables an efficient exploitation of the available resources by continual removal of a prespecified number of unfailed test units at prespecified time points (Type-I PC) or upon observing a failure (Type-II PC). The withdrawn units are typically used in other experiments in the same or at a different facility. Several authors have undertaken an investigation under a PC framework, while dealing with specific parametric life distributions. The book by Balakrishnan and Aggarwala (2000) provides a comprehensive treatment of progressive censoring detailing the theory and application of likelihood based inference under PC. A more recent review by Balakrishnan (2007) documented the research till date on inference based on PC schemes.

Consider n independent and identical items put under life-test, with the lifetimes of each following a continuous distribution with pdf and cdf $f(t;\boldsymbol{\theta})$ and $F(t;\boldsymbol{\theta})$), respectively. Under a Type-I PC scheme, times $T_1 < T_2 < \ldots T_m$ are fixed in advance at which point $R_1, R_2, \ldots, R_m$ units, respectively, are removed from the experiment. The sequence R_i is also fixed apriori. We shall assume throughout the article the feasibility of PC, i.e. the number of units still on test at each censoring time is larger than the corresponding number of units planned to be removed. If k observations are observed to fail during the experimentation at times $x_1, x_2, \ldots, x_k$, respectively, the likelihood function corresponding to this PC scheme can be written as

$$L(\boldsymbol{\theta}) = C \prod_{i=1}^{k} f\left(x_i;\boldsymbol{\theta}\right) \prod_{j=1}^{m} \left[1 - F(T_j;\boldsymbol{\theta})\right]^{R_j}, \quad x_1 < x_2 < \ldots < x_k, \tag{4}$$

where C is a normalizing constant independent of $\boldsymbol{\theta}$. Note that in (4), the number of failures k is a random quantity. Further we assume termination of the entire experiment at T_m in the sense that

$$k + \sum_{j=1}^{m} R_j = n. \tag{5}$$

A Type-II PC obtains when the censoring times match the failure times, and the experiment is terminated after observing a predetermined number of failures. The likelihood of a Type-II PC that is terminated at the m-th failure equals

$$L(\boldsymbol{\theta}) = C^* \prod_{i=1}^{m} f\left(x_i; \boldsymbol{\theta}\right) \left[1 - F(x_i; \boldsymbol{\theta})\right]^{R_i}. \tag{6}$$

The likelihood expression in (6) matches that in (4) by replacing $k = m$. However, note that for the Type-II censoring scheme, the individual censoring times x_i are random, whereas the number of failures m, is a fixed quantity, presenting an exactly reverse scenario than that for Type-I PC. Consequently, the finite-sample inference results are different under the frequentist framework. Note that the Bayesian approach, on the other hand, yields small-sample inference that is identical irrespective of the scheme adopted.

An implicit assumption of progressive censoring scheme is the availability of units to withdraw. This is similar to the restriction imposed on the support of hypergeometric distribution, for example. Strictly speaking, this indeed makes the number of censoring times and number of units to remove random. The likelihood in (4) is thus best interpreted as a likelihood conditional on the R's. With a censoring mechanism that is non-informative of the failure, the usual inference on the basis of (4) remains valid.

The purpose of this article is two fold. Firstly, we consider the Bayesian inference of the two parameters of the LHR distribution for a progressively censored sample. While the prior specification is similar to Lin *et al.* (2006), the treatment is somewhat different in our article. Specifically, we focus on the representation of the posterior distribution as a mixture and make the connection with the frequentist approach more transparent. Also, unlike Lin *et al.* (2006), we cover both Type-I and Type-II censoring in our treatment.

The second aim of this manuscript is to provide methodology for comparing competing progressive sampling schemes with the purpose of obtaining an optimal sampling scheme for a given prior distribution. Optimal censoring schemes have been explored for two criteria: one that is based on minimizing variance of specific percentiles and the other that is derived from minimizing the length of the credible interval for a percentile. Exploration of an optimal censoring scheme is effectively a discrete optimization problem, and is computationally quite expensive. We propose some sub-optimal censoring schemes that are practical from an experimenter's viewpoint and compare them with the complete sample with respect to the chosen criteria.

The rest of the paper is organized as follows. In section 2, we discuss Bayesian inference of the parameters of a LHR under a progressive censoring scheme. Several interesting correspondence with the frequentist framework have been noted. Section 3 details the implementation of the Bayesian methodology for a dataset that has been previously analyzed in a non-Bayesian setup. Optimal Bayesian planning is discussed in section 4 and is supplemented by extensive numerical results. Section 5 concludes with some pertinent discussion.

2 Bayesian Inference

As in Pandey *et al.* (1993) (see also Lin *et al.* 2006), we assume that *apriori* λ_1 and λ_2 are independently distributed as $Gamma(a, b)$ and $Gamma(c, d)$, with pdf given by;

$$\pi_1(\lambda_1) \propto \lambda_1^{a-1} e^{-b\lambda_1}, \qquad \lambda_1 > 0, \tag{7}$$

$$\pi_2(\lambda_2) \propto \lambda_2^{c-1} e^{-d\lambda_2}, \qquad \lambda_2 > 0, \tag{8}$$

respectively. Here the hyper parameters $a, b, c, d > 0$ are assumed known. In the sequel, we adopt the notation corresponding to a Type-I PC. Note, however, that in the Bayesian framework, the statistical inference remains unchanged when we deal with a Type-II PC scheme, or a scheme that adopts a combination of the two. Based on the Type-I progressively censored sample $(\boldsymbol{x}, \boldsymbol{R}, \boldsymbol{T})$, the likelihood function can be written as

$$\begin{aligned} l(data|\lambda_1,\lambda_2) &\propto \prod_{i=1}^{k}(\lambda_1+2\lambda_2 x_i)e^{(-\lambda_1 x_i+\lambda_2 x_i^2)} \times e^{-\sum_{j=1}^{m} R_j(\lambda_1 T_j+\lambda_2 T_j^2)} \\ &\propto \prod_{i=1}^{k}(\lambda_1+2\lambda_2 x_i)e^{-\lambda_1 W_1-\lambda_2 W_2}, \end{aligned} \tag{9}$$

where

$$W_1=\sum_{i=1}^{k} x_i+\sum_{j=1}^{m} R_j T_j, \quad \text{and} \quad W_2=\sum_{i=1}^{k} x_i^2+\sum_{j=1}^{m} R_j T_j^2.$$

In order to accommodate the “no failure” case corresponding to $k=0$, we adopt the convention $\sum_1^0(\cdot)=0$ and $\prod_1^0(\cdot)=1$. Note that in this case MLE's do not exist, while the Bayesian method produces valid estimators.

The joint posterior density function of λ_1 and λ_2 can be written as

$$\begin{aligned} \pi(\lambda_1,\lambda_2|data) &\propto \prod_{i=1}^{k}(\lambda_1+2\lambda_2 x_i)e^{-\lambda_1(b+W_1)}e^{-\lambda_2(d+W_2)}\lambda_1^{a-1}\lambda_2^{c-1} \\ &\propto \left(\lambda_1^k+\lambda_1^{k-1}2\lambda_2\sum_{i=1}^{k}x_i+\lambda_1^{k-2}(2\lambda_2)^2\sum_{1\le i_1<i_2\le k}x_{i_1}x_{i_2}+\ldots+(2\lambda_2)^k\prod_{i=1}^{k}x_i\right)\times \\ &\quad e^{-\lambda_1(b+W_1)}e^{-\lambda_2(d+W_2)}\lambda_1^{a-1}\lambda_2^{c-1} \\ &\propto \sum_{j=0}^{k}\left((2\lambda_2)^j\tau_j\lambda_1^{k-j}\right)e^{-\lambda_1(b+W_1)}e^{-\lambda_2(d+W_2)}\lambda_1^{a-1}\lambda_2^{c-1} \\ &\propto \sum_{j=0}^{k}\tau_j\lambda_1^{a+k-j-1}e^{-\lambda_1(b+W_1)}\quad\lambda_2^{c+j-1}e^{-\lambda_2(d+W_2)}, \end{aligned} \tag{10}$$

where

$$\tau_0=1,\quad \tau_1=\sum_{i=1}^{k}x_i,\quad \tau_2=\sum_{1\le i_1<i_2\le k}x_i x_j\ ,\ldots,\ \text{and}\quad \tau_k=\prod_{i=1}^{k}x_i.$$

Note that (10) can be written as

$$\pi(\lambda_1,\lambda_2|data)=\sum_{j=0}^{k}\pi_j f_{1j}(\lambda_1)f_{2j}(\lambda_2),$$

where for $j = 0, \ldots, k$,

$$f_{1j}(\lambda_1) = Gamma(a + k - j, b + W_1), \quad f_{2j} = Gamma(c + j, d + W_2),$$

$$\pi_j = \frac{p_j}{\sum_{i=0}^{k} p_i}, \quad p_j = \frac{\Gamma(a + k - j)}{(b + W_1)^{a+k-j}} \times \frac{\Gamma(c + j)}{(d + W_2)^{c+j}} \times \tau_j \times 2^j. \tag{11}$$

The marginal posterior density functions of λ_1 and λ_2 are also easily seen to be of the mixture form

$$\pi(\lambda_1|data) = \sum_{j=0}^{k} \pi_j f_{1j}(\lambda_1) \tag{12}$$

and

$$\pi(\lambda_2|data) = \sum_{j=0}^{k} \pi_j f_{2j}(\lambda_2), \tag{13}$$

respectively. The posterior means (Bayes estimates under squared-error loss) are given by

$$E(\lambda_1|data) = \sum_{j=0}^{k} \pi_j \frac{a + k - j}{b + W_1}, \qquad E(\lambda_2|data) = \sum_{j=0}^{k} \pi_j \frac{c + j}{d + W_2} \tag{14}$$

It is interesting to note that the posterior means in (14) satisfy the relation

$$(b + W_1)E(\lambda_1|data) + (d + W_2)E(\lambda_2|data) = a + c + k, \tag{15}$$

which is seen to be an extension of the identity satisfied by the maximum likelihood estimators (MLE). Indeed as all of $a, b, c, d \longrightarrow 0$, thereby yielding the joint noninformative prior $\pi_{ninf}(\lambda_1, \lambda_2) \propto (\lambda_1\lambda_2)^{-1}$, the relation in (15) is identical to that satisfied by the MLE's. Unlike the MLE's however, the Bayes estimators here are in closed-form.

The relation in (15) actually stems from a distributional identity that also has a frequentist counterpart. In order to see that, first note that in view of (11), the joint posterior density of $U = (b + W_1)\lambda_1$ and $V = (d + W_2)\lambda_2$ are given by

$$g(u, v|data) = \sum_{j=0}^{k} \pi_j g_{1j}(u) g_{2j}(v),$$

where g_{1j}, g_{2j} are probability density functions of $Gamma(a+k-j, 1)$, $Gamma(c+j, 1)$, respectively. Let J be a discrete random variable taking values on $\{0, 1, 2, \ldots, k\}$ with $P(J = j) = \pi_j$. Thus, U, V are conditionally independent given $J = j$, with g_{1j}, g_{2j} as respective marginal pdf's. Consequently, conditionally given $J = j$, $U + V$ is distributed as a $Gamma(a + c + k, 1)$ random variable. This distribution, being free of j, is also the unconditional distribution. Thus, *a posteriori*

$$(b + W_1)\lambda_1 + (d + W_2)\lambda_2 \;\sim\; Gamma(a + c + k, 1). \tag{16}$$

The equation in (15) is a trivial consequence of the distributional relation in (16). In the frequentist framework, a result analogous to (16) also holds under Type-II PC, a fact that can be seen by applying the results from Viveros and Balakrishnan (1994) for exponential distribution to data from LHR in the transformed scale $\lambda_1 t + \lambda_2 t^2$.

Lin et al (2006) conduct Bayesian inference and prediction for a more general PC scheme under Type-II censoring. Due to the complex structure of the general estimator, however, they resorted to a numerical comparison between the MLE's and the Bayes estimators.

The distributional result in (16) brings out yet another interesting interpretation of the prior parametrization. In the past decade or so, several researchers have worked on the theoretical properties of *probability matching* priors. A probability matching prior is defined as one that yields the same quantiles for a parametric function of interest *aposteriori* as well as under the frequentist framework upto an order of approximation (c.f. Datta and Sweeting, 2005). In view of (16), one can say that as $a, c \longrightarrow 0$, the joint prior for λ, λ_2 serve as an *exact* probability matching prior for $(b + W_1)\lambda_1 + (d + W_2)\lambda_2$.

The mixture representation of the posterior greatly facilitates the Bayesian inference. One can exploit this structure to obtain simulation based inference wherever needed. For example, percentiles from the marginal posterior distributions, while not available in closed form, can easily be estimated from samples generated from the corresponding posterior using Gibbs sampling. The credible intervals for the individual parameters λ_1 and λ_2 can be formed using these sample-based percentiles. Note, however, that due to the multimodal nature of the densities, the usual procedure for finding highest posterior density (HPD) intervals will not be appropriate here. Next we indicate an procedure to obtain a joint credible set for the parameters of LHR that has an exact pre-specified probability content.

2.1 Construction of a Joint Credible Set

The joint credible set for the parameters of LHR are based on the posterior distributions of certain quantities. To fix ideas, let us fix $1 \leq m_1 < m$, and define the variables

$$\begin{aligned} B_1 &= 2\lambda_1(b + W_1) + 2\lambda_2(d + W_2) \\ B_2 &= \frac{\lambda_1(b + \widetilde{W}_1) + \lambda_2(d + \widetilde{W}_2)}{\lambda_1(b + W_1) + \lambda_2(d + W_2)}, \end{aligned} \qquad (17)$$

where

$$\widetilde{W}_1 = \sum_{i=1}^{k_1} x_i + \sum_{j=1}^{m_1} R_j T_j, \quad \widetilde{W}_2 = \sum_{i=1}^{k_1} x_i^2 + \sum_{j=1}^{m_1} R_j T_j^2, \quad 1 \leq k_1 \leq k.$$

Clearly, $\widetilde{W}_1, \widetilde{W}_2$ are versions of W_1, W_2 when only considers the observations up-to T_{m_1}. Note that under such a truncation, k_1, the number of observed failures, which is still random, can be less than or equal to k. The procedure below is developed for any $m_1 < m$. We shall discuss the issue of choosing a suitable m_1 later. The following lemma provides the crucial result in developing the joint credible set.

Lemma 2.1: A posteriori, the random quantities B_1 and B_2 are independent, with B_1 distributed as a chi-square with 2(a+c+k) degrees of freedom and B_2 distributed as mixture of beta random variables with pdf $g_2(b_2) = \sum_{j=0}^{k} \pi_j Beta_{[\beta_1,\beta_2]}(a + k - j, c + j)$, with $\beta_1 = \min\{(b+\widetilde{W}_1)/(b+W_1), (d+\widetilde{W}_2)/(d+W_2)\}$, $\beta_2 = \max\{(b+\widetilde{W}_1)/(b+W_1), (d+\widetilde{W}_2)/(d+W_2)\}$, and $Beta_{[\beta_1,\beta_2]}(\alpha_1, \alpha_2)$ is the pdf of a beta random variable given by

$$q(y) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cdot \frac{(y-\beta_1)^{\alpha_1-1}(\beta_2-y)^{\alpha_2-1}}{(\beta_2-\beta_1)^{\alpha_1+\alpha_2-1}}.$$

The proof is a straightforward application of transformation of variables using (17) and the joint posterior pdf of (λ_1, λ_2), and is omitted. For any $0 < \alpha < 1$, let $\gamma = 1 - \sqrt{1-\alpha}$, and let $b_1^{(1)}$, $b_1^{(2)}$ denote the $100(\gamma/2)$th and $100(1-\gamma/2)$th percentiles of a chi-square distribution with $2(a+c+k)$ degrees of freedom. Furthermore, if $b_2^{(1)}$ and $b_2^{(2)}$ denote the corresponding percentiles of the posterior distribution of B_2, then, exploiting the independence of B_1, B_2 from Lemma 2.1, the set

$$\mathcal{C}(\lambda_1,\lambda_2) = \Big\{(\lambda_1,\lambda_2):\ b_1^{(1)} < B_1 < b_1^{(2)},\ \ b_2^{(1)} < B_2 < b_2^{(2)}\Big\} \tag{18}$$

has exact probability content $1-\alpha$.

To investigate the form of the confidence set $\mathcal{C}(\lambda_1, \lambda_2)$, we note that as a function of λ_1, λ_2, B_2 only varies with the ratio λ_2/λ_1 and is monotonically increasing(decreasing) in the same if $(b+\widetilde{W}_1)/(b+W_1) < (>) (d+\widetilde{W}_2)/(d+W_2)$. Therefore, $b_2^{(1)} < B_2 < b_2^{(2)}$ is realized if and only if the ratio λ_2/λ_1 is contained in an interval with end points A_1 and A_2 where

$$A_1 = \frac{(b+\widetilde{W}_1) - b_2^{(2)}(b+W_1)}{b_2^{(2)}(d+W_2) - (d+\widetilde{W}_2)}, \quad A_2 = \frac{(b+\widetilde{W}_1) - b_2^{(1)}(b+W_1)}{b_2^{(1)}(d+W_2) - (d+\widetilde{W}_2)}.$$

Since the percentiles $b_2^{(1)}, b_2^{(2)}$ come from a mixture of beta pdf's each with support on $\big(\min\{(b+\widetilde{W}_1)/(b+W_1), (d+\widetilde{W}_2)/(d+W_2)\},\ \max\{(b+\widetilde{W}_1)/(b+W_1), (d+\widetilde{W}_2)/(d+W_2)\}\big)$, they also lie within these bounds. Consequently, the ratio λ_2/λ_1 lies within $(\min\{A_1, A_2\}, \max\{A_1, A_2\})$ with (posterior) probability $1-\gamma$. We can thus express the joint credible set in (18) in a more explicit form:

$$\begin{aligned}\mathcal{C}(\lambda_1,\lambda_2) = \Big\{(\lambda_1,\lambda_2):\ & b_1^{(1)}/2 < \lambda_1(b+W_1)+\lambda_2(d+W_2) < b_1^{(2)}/2, \\ & \lambda_1\cdot\min\{A_1,A_2\} < \lambda_2 < \lambda_1\cdot\max\{A_1,A_2\}\Big\}.\end{aligned} \tag{19}$$

The set in (19) has the simple structure of a trapezoid bounded on the positive quadrant in the $\lambda_1 : \lambda_2$ space by the parallel lines

$$\lambda_1(b + W_1) + \lambda_2(d + W_2) = b_1^{(1)}/2, \quad \lambda_1(b + W_1) + \lambda_2(d + W_2) = b_1^{(2)}/2, \tag{20}$$

and the two lines through origin

$$\lambda_2 = A_1\lambda_1, \quad \lambda_2 = A_2\lambda_1. \tag{21}$$

One reasonable choice for m_1 then is the one that minimizes the area of this trapezoid. Note that m_1 plays a role only in constructing A_1, A_2, and hence in the location of the lines given by (21). For a given progressively censored sample and prior model, the location of the parallel lines in (20) is completely specified and does not vary with the choice of m_1. It thus follows, after some algebra, that minimizing the area of the trapezoid tantamount to minimizing the distance criterion

$$C(m_1) = \left| \frac{1}{(d + W_2)A_1 + (b + W_1)} - \frac{1}{(d + W_2)A_2 + (b + W_1)} \right| \tag{22}$$

as a function of the percentiles of B_2, the beta-mixture random variable.

2.2 Prediction

In the context of observing progressively censored samples, the prediction problem typically centers around predicting the times to failure of the *last* R_m units surviving at the time of termination of the experiment. In our context, this translates to predicting $x_{k+1}, x_{k+2}, \ldots, x_{R_m}$. Note that $x_{k+l},\ 1 \leq l \leq R_m - k$, the l-th smallest of the last R_m failure times can be regarded as the last failure time from a progressively censored sample with censoring schemes $(T_1, R_1), (T_2, R_2), \ldots, (T_{m-1}, R_{m-1}), (T_m, 0), (x_{k+l}, R_m - l)$ and failure times at $x_1 < x_2 < \ldots < x_k < x_{k+1} \ldots < x_{k+l}$. Note thus that the prediction framework is that of a mixed Type-I and Type-II progressively censored scheme. The marginal predictive distribution of x_{k+l} can be derived as

$$\pi(x_{k+l}|data) = \int\int \pi(x_{k+l}|data; \lambda_1, \lambda_2)\pi(\lambda_1, \lambda_2|data)d\lambda_1 d\lambda_2. \qquad (23)$$

Equation (23) has an explicit expression and extends the results in Lin et. al. (2006) derived under Type-II progressive censoring. The distribution, however, is complex, non-standard, and does not lend itself to easy inference. Instead of repeating the endeavor of Lin et. al. (2006), we present a simulation based approach using Gibbs sampling, where samples in each step can be generated from distributions that are tractable. Further, in this method, we generate the joint distribution of any subset of $x_{k+1}, x_{k+2}, \ldots, x_{R_m}$, in one stroke. Since the approach is based on simulation, quantities of interest based on the predictive distribution can be estimated as an appropriate function of the drawn samples.

To fix ideas, let us concentrate on estimating the predictive distribution of x_{k+l}. We shall augment the data with samples from the distribution of the latent variables $x_{k+1}, x_{k+2}, \ldots, x_{k+l-1}$. This will facilitate the generation of tractable full conditionals using the Markovian nature of the progressively censored samples. The Gibbs sampling steps consist of generation of two sets of full conditionals:

Failure Times: Conditionally given the data, and λ_1, λ_2, the ensemble $\{x_{k+1}, \ldots, x_{k+l}\}$ is distributed as the first l order statistics from a random sample of size R_m from a $LHR(\lambda_1, \lambda_2)$ left truncated at T_m. Note that since the LHR cdf F has an explicit inverse function

$$F^{-1}(t) = \frac{-\lambda_1 + \sqrt{\lambda_1^2 - 4\lambda_2 \ln(1-t)}}{2\lambda_2}, \quad 0 < t < 1,$$

a random draw t from the left truncated LHR can be made easily by

(a) drawing a random variate u from $uniform(0, 1)$,

followed by

(b) calculating $t = F^{-1}\big(F(T_m) + \big(1 - F(T_m)\big) \cdot u\big)$.

Model Parameters: The full conditional distribution of (λ_1, λ_2) given the rest has the form of a $(k+l+1)$-component mixture of independent copies of $Gamma(a+k+l-j, b+W_1^*)$, and $Gamma(c+j, d+W_2^*)$, where

$$W_1^* = \sum_{i=1}^{k+l} x_i + \sum_{j=1}^{m-1} R_j T_j + (R_m - l)x_{k+l}, \quad W_2^* = \sum_{i=1}^{k+l} x_i^2 + \sum_{j=1}^{m-1} R_j T_j^2 + (R_m - l)x_{k+l}^2.$$

This follows easily following the same line of derivation used in establishing the joint posterior for (λ_1, λ_2). The mixing proportions have the same form as (11) with k, W_1, W_2 replaced by $k+l, W_1^*, W_2^*$, respectively. The mixture structure imposes an additional data augmentation step involving the latent discrete random variable with probability mass function given by the mixing proportions. Note however, that in each step of the simulation we are generating random variates from known distributions which make the implementation easy and fast.

3 An Example

Our methodology is illustrated on the failure data from a life test performed on prototypes of a small electrical appliance. The data which was originally presented in Nelson (1970), consists of number of cycles to failure or censoring of 36 units. One complicating feature of the data that we shall ignore for the present analysis is the existence of multiple failure modes. A Type-II progressively censored sample with 10 failures ($k = 10$) and censoring scheme $R_1 = \cdots = R_9 = 2, R_{10} = 8$ was randomly generated from the data and analyzed by Kundu (2008) using a Weibull distribution. The ordered failure times from the progressively censored sample are: 11, 35, 49, 170, 329, 958, 1925, 2223, 2400 and 2568. For computational stability, we re-express the time units in hundreds of cycles.

The maximum likelihood estimate of λ_1 using a $LHR(\lambda_1, \lambda_2)$ turns out to be 0.022. On the other hand it appears that the profile likelihood of λ_2 is maximized at or near the boundary of the parameter space which makes it difficult to base the inference on maximum likelihood. A similar feature was also reflected in the interval estimates of the scale parameter of the Weibull distribution spilling over to the negative real axis (Kundu, 2008). Bayesian analysis, by focusing on the average of the posterior likelihood, avoids such difficulties.

Three sets of gamma priors are chosen for λ_1, λ_2 with parameters a, b, c, d all equaling 1, 2, 5, respectively. The prior mean equals 1 in each case, while the prior variance equals 1, 0.5, 0.2 representing a gradual decrease in prior uncertainty i.e. an increase in the degree of belief in the prior formulation. Figure 1 demonstrates the effect of prior information on the posterior distributions of λ_1, λ_2. The posterior seems sensitive to the choice of prior parametrization, specifically for λ_2. It seems that a stronger prior information leads to a more pronounced uni-modal character. The shift of the posterior mass as a function of the prior parameter is also evident from the figure.

Table 2 provides numerical summary measures from the posterior based on the different prior specification. While the posterior mean and median are sensitive to the prior choice, often changing two- or four-fold across different sets of prior parameters, the variability remain quite stable for both λ_1 and λ_2. A stronger prior information has a stronger symmetrizing effect on the posterior, as is evident both from Figure 1 and the credible intervals reported in Table 2.

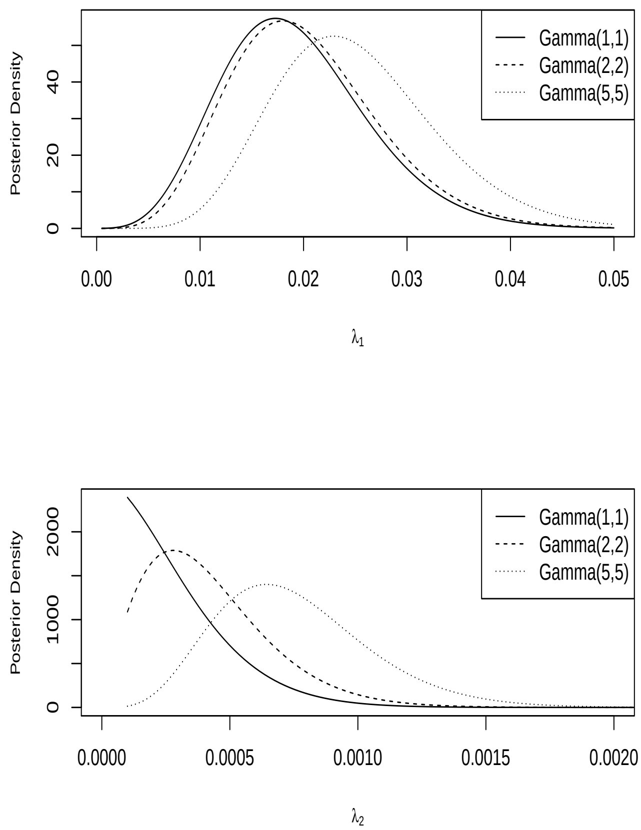

Figure 1: Posterior Distribution of λ_1, λ_2 for priors: Gamma(1,1) (solid), Gamma(2,2) (dashed), Gamma(5,5) (dotted).

Table 2: Posterior summary measures for λ_1, λ_2 under different prior parametrization

	λ_1				λ_2			
Prior	Mean	Median	SD	95% CI	Mean	Median	SD	95% CI
$\lambda_1, \lambda_2 \sim Gam(1,1)$	0.018	0.017	0.007	(0.007, 0.034)	0.0003	0.0003	0.0002	(0.00002,0.0009)
$\lambda_1, \lambda_2 \sim Gam(2,2)$	0.02	0.019	0.0073	(0.037, 0.008)	0.0004	0.00037	0.0002	(0.00007,0.001)
$\lambda_1, \lambda_2 \sim Gam(5,5)$	0.027	0.026	0.008	(0.013, 0.044)	0.0007	0.00065	0.0003	(0.0002,0.0014)

The contour plot of the joint posterior (11) of (λ_1, λ_2) in Figure 2 is drawn for a $Gamma(2,2)$ prior for both λ_1 and λ_2. It clearly shows the existence of a joint posterior mode that can be visually read of the plot as $\tilde{\lambda}_1 \approx 0.02$, $\tilde{\lambda}_2 \approx 0.0003$ matching the posterior means in Table 2. The plot further demonstrates a negative association between λ_1, λ_2. Indeed the correlation is estimated to be -0.265 based on a random draw of 500 pairs of (λ_1, λ_2) from the joint posterior. Plot of the posterior mean survival function (Figure 3) for this model exhibits increasing uncertainty with increase in time.

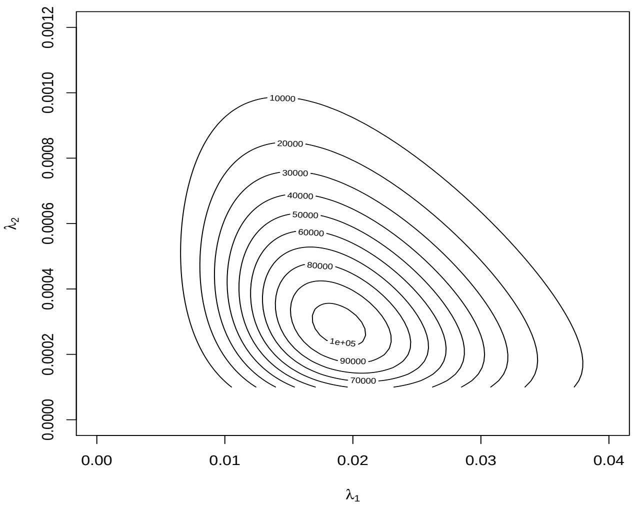

Figure 2: Contour plot of the joint posterior Distribution of (λ_1, λ_2) for Gamma(2,2) prior.

Construction of a joint credible set for λ_1, λ_2 has been pursued using the method described in Section 2.1. The criterion function $C(m_1)$ in (22) was numerically minimized at $m_1 = 5$, which yielded a confidence set

$$\mathcal{C}(\lambda_1, \lambda_2) \;=\; \{(\lambda_1, \lambda_2):\; 15.31 < B_1 < 44.46, \quad 0.014 < B_2 < 0.038\}$$

that has nominal probability content of 90%. Figure 4 indicates this joint set as the area enclosed by the trapezoid. Note that although $C(m_1)$ was minimized at $m_1 = 5$, the differences in the C function for different m_1 is minor, indicating that any choice of m_1 would be equally efficient. The estimated coverage probability of the confidence set based on a random sample of 50000 pairs of (λ_1, λ_2) from (11) for all m_1 varied between 86–87%.

In the Bayesian context, a widely used model selection tool is the *Deviance Information*

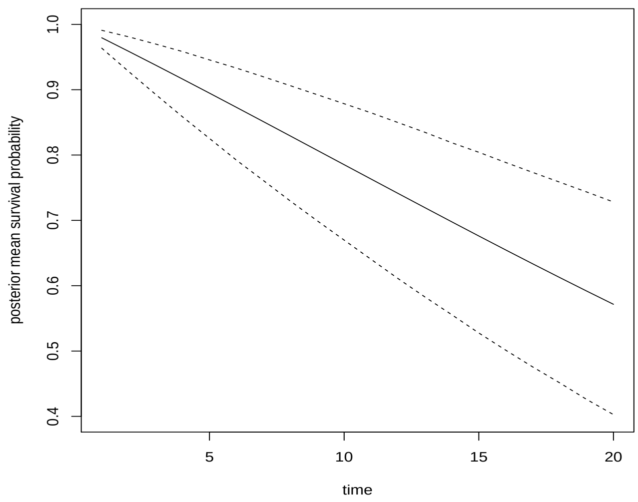

Figure 3: Survival function estimate (solid) with point-wise 95% credible bounds (dashed).

Criterion (DIC) popularized by Spiegelhalter *et al.* (2002). The attractiveness of DIC lie in its ease of computation and its applicability across general sets of models, nested or non-nested. Apart from some constant that remains unchanged between models, the comparison of DIC translates into comparing the quantity

$$2L\left(E_{post}(\boldsymbol{\theta})\right) - 4E_{post}\left(L(\boldsymbol{\theta})\right), \tag{24}$$

where $L(\boldsymbol{\theta})$ is the log-likelihood evaluated at $\boldsymbol{\theta}$, and E_{post} denotes the posterior expectation. Model with the minimum DIC is preferred over the others. Table 3 provides the DIC values for LHR, the two limiting models, namely, Exponential and Rayleigh as well as Weibull with parameters α and β, with $Gamma(2,2)$ prior chosen for all model parameters. In evaluating (24), for all but Weibull, the first term is easily calculated based on the closed-form posterior mean of the model parameters. For Weibull, we take recourse to Markov Chain Monte Carlo (MCMC) technique, which comprises of a Gibbs sampling step for generation of α, and a weighted bootstrap step (see, for example, Smith and Gelfand 1992) for the generation of β. Posterior mean of the log-likelihood is estimated as an average of 5000 evaluations of the log-likelihood based on samples from the posterior. The magnitude of DIC favors Weibull model

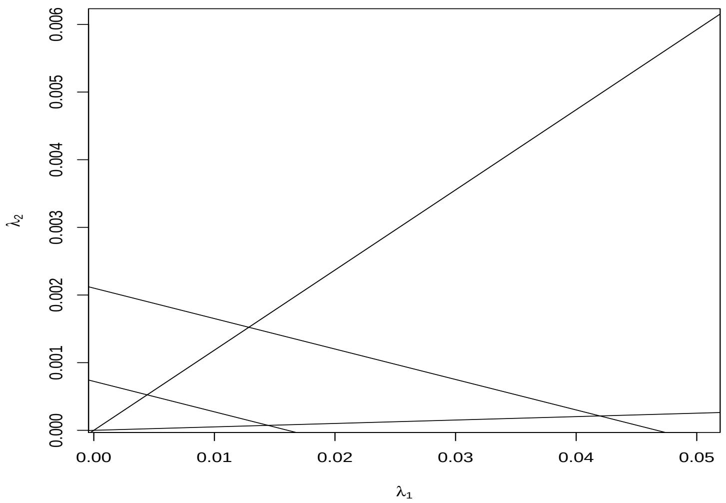

Figure 4: Joint credible set for (λ_1, λ_2) for $m_1 = 5$.

over the others. However, as Spiegelhalter *et al.* (2002) points out, a difference in DIC larger than 7 establishes clear superiority of one model over the other, while a DC between 3 and 7 should not be used as a sole comparator between models. In our case, indeed the Weibull, Exponential and LHR DIC values are all within a maximum of 4 points of each other. Table 3 also provides posterior mean of median reliability (B_{50}) along with the 500- and 2000-cycle reliability as comparative metrics across models. Weibull seems to stand out from the remaining models in its strikingly higher point estimate of median reliability. The Weibull model further demonstrates a large uncertainty in the estimate as reflected through the credible interval. The 500-cycle reliability across models also demonstrate some differences. In particular, the estimates for Weibull are markedly smaller compared to the other models. The Kaplan-Meier based nonparametric estimates and associated 95% confidence intervals for the 500-cycle reliabilities turn out to be 0.84 (0.71, 0.97). On the other hand, the models are quite comparable when estimating the reliability at 2000 cycles. The corresponding nonparametric point and interval estimates are 0.76 and (0.60, 0.92), somewhat higher than the model-based estimates. Based on the DIC, the Rayleigh model seems to be a much poorer fit compared to the others.

Table 3: Comparative Summary of Different Models

Sampling Distribution	Prior	DIC	B_{50} (CI)	Reliability (CI)	
				500-cycle	2000-cycle
$LHR(\lambda_1, \lambda_2)$	$\lambda_1, \lambda_2 \sim Gamma(2, 2)$	101.58	24.42 (16.20, 37.21)	0.90 (0.83, 0.95)	0.57 (0.40, 0.74)
$Exponential(\lambda_1)$	$\lambda_1 \sim Gamma(2, 2)$	99.16	30.04 (17.06, 53.75)	0.88 (0.82, 0.94)	0.61 (0.44, 0.77)
$Rayleigh(\lambda_2)$	$\lambda_2 \sim Gamma(2, 2)$	122.47	25.44 (19.25, 34.10)	0.97 (0.95, 0.99)	0.64 (0.47, 0.79)
$Weibull(\alpha, \beta)$	$\alpha, \beta \sim Gamma(2, 2)$	97.66	76.92 (16.3, 247.8)	0.80 (0.67, 0.90)	0.63 (0.46, 0.78)

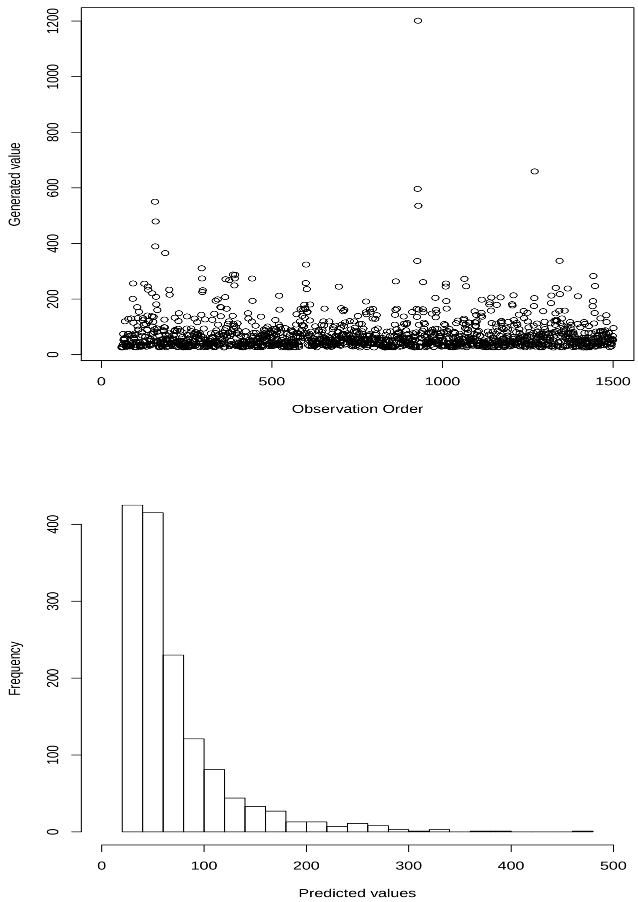

Figure 5: Top Panel: Trace plot for the MCMC samples from the predictive distribution for the first future failure time, Bottom Panel: Histogram of the same samples.

With the priors for λ_1, λ_2 both specified as $Gamma(2, 2)$, the MCMC steps described in Section 2.2 have been followed to generate the prediction distribution for the future failure times for the eight items censored at the last observation time. A sample of size 2000 is drawn using the Gibbs sampling steps, from which the last 1500 have been used for summary estimates of the prediction distribution. The top panel of Figure 4 shows a trace plot of these 1500 observations from the prediction distribution of the *first* future time-of-failure random variable. The flatness of the plot is indicative of stability of convergence. Histogram of the prediction distribution is presented in the bottom panel of the figure. The prediction distribution is highly right skewed with a predicted mean and median time of failure equaling 7180 and 5330 cycles, respectively. The estimated standard deviation of the prediction distribution is 6460 cycles. The extent of skewness in the prediction distribution increases as one moves from a lower to a higher rank-order of future time-of-failure random variables.

4 Planning an Optimal Progressive Censoring Scheme

In an engineering context, a very important aspect of a life-testing experiment is planning an optimal data collection strategy. In this section, we investigate the issues in planning an optimally progressive censored sample collection where the unit failures are governed by an underlying LHR distribution.

From a planning perspective, the possible censoring schemes to be considered are different for Type-I and Type-II progressive censoring schemes. For simplicity of exposition, in this article we restrict ourselves to the exploration of optimal censoring strategy for Type-II censoring. For this, we assume n, the number of units on test and k, the number of failures, to be fixed in advance. Under this framework, two competing Type-II progressive censoring schemes, say $P_1 = \{R_1^{(1)}, \ldots, R_k^{(1)}\}$ and $P_2 = \{R_1^{(2)}, \ldots, R_k^{(2)}\}$ should both satisfy $\sum_{i=1}^k R_i^{(1)} = \sum_{i=1}^k R_i^{(2)} = n - k$.

4.1 Precision Criteria

One can choose from a variety of criteria for choosing an optimal sampling design. Traditionally, the most commonly used among these are the ones dealing directly or indirectly with some dispersion measure. One reasonable criterion proposed by Zhang and Meeker (2005), and further investigated by Kundu (2008) in the Bayesian context of planning a censored data collection strategy for Weibull distribution, is based on minimizing the posterior variance of a percentile of the underlying life-distribution. This will be our first chosen criterion for comparison. The p-th quantile $(0 < p < 1)$ of the LHR distribution is

$$\psi_p = \frac{-\lambda_1 + \sqrt{\lambda_1^2 - 4\lambda_2 \ln(1-p)}}{2\lambda_1}. \tag{25}$$

For a given Type-II progressively censored sample P from LHR, the criterion we want to minimize is given by

$$C_P^{(1)}(p) = E_{data}\{Var_{posterior(P)}(\psi_p)\} \tag{26}$$

where P denotes the reliability plan and $Var_{posterior(P)}(\psi_p)$ denotes the posterior variance of ψ_p for the given plan. Also, E_{data} presents the average over the marginal distribution of the data, which makes (26) a legitimate planing criterion.

A second criterion we consider in this article is the average length of the 95% credible interval for the p-th percentile given by

$$C_P^{(2)}(p) = E_{data}\left(q_{0.975}(\psi_p) - q_{0.025}(\psi_p)\right), \tag{27}$$

where q_δ is the 100δ-th percentile of the posterior distribution of ψ_p. The optimal plan P is the one that minimizes (27).

Clearly, both (26) and (27) are functions of a specific percentile, choice of which may potentially drive the search for the optimal plan. If the experimenter is simultaneously interested in multiple percentiles or a range of them, then it seems natural to set up a criterion that covers them simultaneously. In that vein, we also investigate objective functions

$$\overline{C}^{(1)} = E_{data}\left[\int_0^1 Var_{posterior(P)}(\psi_p)dw(p)\right], \ \overline{C}^{(2)} = E_{data}\left[\int_0^1 \left(q_{0.975}(\psi_p) - q_{0.025}(\psi_p)\right) dw(p)\right], \tag{28}$$

where $0 \leq w(p) \leq 1$, to be specified by the experimenter apriori, is a non-negative weight function on $[0, 1]$ that reflects the degree of importance on estimation of a certain percentile.

4.2 Numerical Findings

We carried out an extensive numerical search for an optimal progressive censoring strategy for a range of fixed n and k values. For each plan we computed (26)–(28). Since the expressions in (26)–(28) are not closed-form, we take recourse to a simulation based approach that is easy to implement. We present below the details of the computational steps.

Computational Algorithm

Step 1: Generate λ_1 and λ_2 from the joint prior distribution of λ_1 and λ_2.

Step 2: Generate a Type-II progressively censored sample $x_1 < \ldots < x_k$ from the given censoring scheme P, under LHR(λ_1, λ_2) assumption for the life-distribution. Since LHR has a closed-form expression for percentiles, this step was easily implemented using the efficient sample generation algorithm recommended by Balakrishnan and Aggarwala (2000), for which the knowledge of R_i's is sufficient.

Step 3: Both $Var_{posterior(P)}(\psi_p)$ and $(q_{0.975}(\psi_p) - q_{0.025}(\psi_p))$ as well as their weighted averages in (28) were calculated on the basis of M_1 pairs of (λ_1, λ_2) sampled from (11).

Step 4: Repeat Steps 1-3, say M_2 times and calculate the average values of the measures obtained in Step 3 which serve as approximations to $C_P^{(1)}, C_P^{(2)}, \overline{C}^{(1)}$ and $\overline{C}^{(2)}$.

For each set of n and k values, our investigation consisted of a search over all possible PC configurations utilizing steps 1–4, with both M_1 and M_2 set at 1000. Different sets of prior parameterizations have also been studied. Instead of overwhelming the readers with the vast amount of information, we present in Table 4 a snapshot of the simulation results that typifies the findings. For each pair of (n, k), we present the optimal configurations along with some practical configurations such as the traditional Type-II, *reverse* Type-II that censors all planned to be censored at the first failure, all censoring in the *middle*, and the *uniform* censoring consisting of balanced withdrawals throughout the observation process. In order to evaluate (28), we used the uniform weight function $w(p) = 1$. In the table, the notation a*b corresponding to a configuration denotes a withdrawals at b successive failure times. The table entries express the reciprocal of the criterion value for each configuration relative to that for the complete sample, so that the resulting ratio is between 0 and 1 and represents the *relative efficiency* of the configuration. For illustrative purposes, we only report the findings for $p = 0.1, 0.5, 0.9$. The prior parameters for λ_1, λ_2 are all set to be equal to 2. In addition to the relative efficiencies, the table also reports for each configuration the average *total time on test* represented by W_1 in this article. Below we summarize the highlights of our findings.

Table 4: Relative efficiency using criteria $C_P^{(1)}, C_P^{(2)}, \overline{C}^{(1)}, \overline{C}^{(2)}$ for different percentiles, n and k

		Configuration	Variance				Credible Interval				
n	k	$(R_1, R_2, \ldots, R_k)$	$p = 0.1$	$p = 0.5$	$p = 0.9$	Ave.	$p = 0.1$	$p = 0.5$	$p = 0.9$	Ave.	$E(W_1)$
15	5	(0*4,10)	0.87	0.81	0.75	0.80	0.89	0.86	0.84	0.87	3.42
		(10,0*4)	0.83	0.86	0.91	0.84	0.80	0.83	0.89	0.84	4.56
		(0*2,10,0*2)	0.84	0.85	0.80	0.82	0.83	0.84	0.86	0.84	3.87
		(2*5)	0.83	0.85	0.84	0.83	0.84	0.83	0.85	0.85	3.91
	10	(0*9,5)	0.94	0.92	0.89	0.91	0.94	0.92	0.88	0.90	3.73
		(5,0*9)	0.85	0.89	0.95	0.87	0.85	0.90	0.95	0.88	4.83
		(0*4,5,0*5)	0.89	0.90	0.93	0.87	0.88	0.90	0.92	0.90	4.23
		(0*2,1*5,0*3)	0.90	0.91	0.93	0.88	0.87	0.89	0.90	0.91	4.35
20	5	(0*4,15)	0.75	0.73	0.68	0.72	0.77	0.74	0.71	0.69	3.69
		(15,0*4)	0.64	0.69	0.72	0.68	0.65	0.72	0.75	0.70	4.72
		(0*2,15,0*2)	0.68	0.70	0.71	0.69	0.67	0.72	0.72	0.69	3.98
		(3*5)	0.69	0.71	0.71	0.71	0.67	0.72	0.71	0.71	3.88
	10	(0*9,10)	0.85	0.80	0.79	0.81	0.86	0.79	0.73	0.81	4.53
		(10,0*9)	0.81	0.84	0.88	0.82	0.78	0.82	0.86	0.83	5.21
		(0*4,10,0*5)	0.81	0.81	0.83	0.82	0.81	0.82	0.83	0.83	5.11
		(1*10)	0.83	0.81	0.82	0.81	0.82	0.81	0.81	0.82	5.01
30	5	(0*4,25)	0.68	0.61	0.56	0.59	0.69	0.62	0.59	0.64	4.12
		(25,0*4)	0.54	0.60	0.69	0.61	0.55	0.61	0.68	0.60	4.95
		(0*2,25,0*2)	0.62	0.63	0.61	0.63	0.61	0.61	0.62	0.63	4.55
		(5*5)	0.69	0.71	0.71	0.71	0.67	0.72	0.71	0.71	4.63
	10	(0*9,20)	0.78	0.73	0.64	0.71	0.79	0.74	0.67	0.73	4.53
		(20,0*9)	0.66	0.69	0.75	0.71	0.68	0.71	0.76	0.70	5.21
		(0*4,20,0*5)	0.72	0.71	0.74	0.73	0.73	0.72	0.72	0.73	4.75
		(2*10)	0.71	0.70	0.71	0.72	0.73	0.71	0.70	0.70	4.93

Summary of Findings:

- The relative efficiencies for either criterion increase as the number of failures increase for a given number of test units n. The efficiencies drop rapidly and become quite low as $n - k$ grows for a fixed k and a given percentile.

- Among the configurations considered, the regular Type-II (censor all at the end) scheme generally seems to perform the best with respect to either the variance or the credible interval length criterion for low percentiles. For these, the worst performance is obtained from the *reverse* Type-II ($R_1 = n - k$) scheme in general. For estimation in the center ($p = 0.5$), however, the efficiencies are quite close to each other and there is no clear-cut winner. On the other hand, the *reverse* Type-II ($R_1 = n - k$) censoring scheme seems to have a slight edge over the others when the interest lies at extremely high percentiles with the regular Type-II generally performing the worst here.

- For the regular Type-II censoring scheme, efficiency clearly decreases with an increase in percentile with the opposite occurring for the reverse Type-II. This feature is evident for all n, k pairs. For either the *uniform* censoring scheme or the scheme censoring in the middle, the efficiency stays about the same for the entire range of percentiles considered.

These basic features remain as we move from one set of parameterizations to another. The computational burden, for any parameterization, increases significantly as n, the number of test units increase due to the increase in the number of possible configurations to search over. One reasonable approach in this case is to restrict the search over a class of a limited number of configurations that are deemed practical and easy to implement.

5 Discussion

In this article, Bayesian inference and planning for linear hazard rate distribution are investigated in detail under a progressively censored sampling scheme. The competing risks structure and existence of simple physical interpretation have attracted investigation of LHR as a potential life-distribution for many years. The main purpose of this article is to take a Bayesian look ar the statistical inference related to LHR under a PC scheme. One obvious advantage of the Bayesian framework is the ability to unify the inference results for both Type I and Type II and even a mix of the two. The development can be generalized to the PC scheme that starts the observation process at the r-th order statistics ($r > 1$), albeit at the expense of tractable inference. We did not pursue that route here and instead focused on the case that is perhaps more practical and brings forth some interesting connection with the frequentist treatment.

Krane (1963) studied hazard rates in the general polynomial form

$$h_1(t) = \beta_0 + \beta_1 t + \cdots + \beta_m t^m, \quad m \geq 1,$$

that includes, apart from LHR, non-monotonic bathtub-shaped hazard functions. The development in this article readily generalizes to the polynomial hazard model under independent gamma priors for the β parameters. For a given application, however, a higher order than a third degree polynomial would rarely be of use.

Throughout this article, we have worked with the parameter space consisting of the positive quadrant in two dimensions. This makes either Exponential or Rayleigh limiting cases of our LHR formulation. An alternative formulation that includes these distributions as special cases of LHR adds on to the parameter space the boundary $B = \{(\lambda_1 = 0, \lambda_2 > 0) \cup (\lambda_1 > 0, \lambda_2 = 0)\}$. The large sample inference results of the maximum likelihood estimators of $(\lambda_1, \lambda_2) \in B$ are non-standard. Using some general theory advanced by Self and

Liang (1987), one can show for regular Type-II censored data that the MLE's of both λ_1, λ_2 converge in distribution to a 50-50 mixture of a normal random variable and a degenerate random variable at 0 (c.f. Sen 2006). In the Bayesian framework, the modified parameter space is accommodated by means of a mixture prior that can be defined hierarchically as

$$\begin{aligned} \pi^*(\lambda_1, \lambda_2 \mid w_1, w_2) &= w_1 \pi_1(\lambda_1) \delta_{\{\lambda_2=0\}} + w_2 \delta_{\{\lambda_1=0\}} \pi_2(\lambda_2) + (1 - w_1 - w_2) \pi_1(\lambda_1) \pi_2(\lambda_2) \\ \tilde{\pi}(w_1, w_2) &\propto w_1^{\alpha_1 - 1} w_2^{\alpha_2 - 1} (1 - w_1 - w_2)^{\alpha_3 - 1}, \quad \alpha_1, \alpha_2, \alpha_3 > 0, \end{aligned} \tag{29}$$

where $\pi_1(\lambda_1), \pi_2(\lambda_2)$ conform to the gamma priors in (7) and (8), respectively, and $\delta_{\{\}}$ refers to the measure degenerate at zero. The prior specification incorporates the two boundaries of the parameter space. The mixture prior does not add any additional complexity in the Bayesian framework when the inference is carried out numerically using MCMC. We did not pursue this in this article, however, in order to keep the exposition simple.

The planning component described in this article constitutes an important practical aspect of a life-testing experiment. Both criteria explored here relate to precision under the Bayesian framework. In choosing the prior parameterization, the experimenter should perhaps rely on historical data. Optimal planning is useful in view of the fact that an experimenter planning the experiment has several PC configurations to choose from, the number of which increases quickly as n increases. If freeing up units to make them available for another experiment is a primary goal, then understanding the optimality (or otherwise) of a PC scheme that censors earlier in the process is particularly crucial. It seems that for LHR, both the variance and the credible interval criteria indeed find such PC schemes to have desirable properties, especially when the number of available testing units is small, and interest lies on the larger percentiles of the life distribution. The numerical search strategy for finding optimum PC can be prohibitively time consuming in general. Further research is needed to find an efficient search algorithm in this regard.

References

[1] Ashour, S.K. and Youssef, A. (1991), "Bayes estimation of the linear failure rate", *Journal of Indian Association of Productivity Quality and Reliability*, **16**, 9 - 16.

[2] Bain, L.J. (1974), "Analysis for the linear failure rate life testing distribution", *Technometrics*, **16**, 551 - 559.

[3] Balakrishnan, N. (2007), "Progressive censoring methodology: An apprisal", (with discussions), *Test*, **16**, 211–289.

[4] Balakrishnan, N. and Aggarwala, R. (2000), *Progressive censoring: theory, methods, and applications* Boston: Birkhuser.

[5] Balakrishnan, N. and Malik, H.J. (1986), "Order statistics from the linear failure rate distribution, Part I: Increasing Hazard Rate Case", *Communications in Statistics - Theory and methods*, **15**, 179 - 203.

[6] Boag, J. W. (1949), "Maximum likelihood estimates of the proportion of patients cured by cancer therapy", *Journal of the Royal Statistical Society, Series B*, **11**, 15–44.

[7] Broadbent, S. (1958), "Simple mortality rates", *Journal of the Royal Statistical Society, Series C*, **7**, 86–95.

[8] Carbone, P. O., Kellerhouse, L. E., and Gehan, E. A. (1967), "Plasmacytic Myeloma: a study of the relationship of survival to various clinical manifestations and anomalous protein type in 112 patients", *The American Journal of Medicine* **42**, 937–948.

[9] Cohen, Jr., A. C. (1963), "Progressively Censored Samples in Life Testing," *Technometrics* **5**, 327–339.

[10] Datta, G. S. and Sweeting, T. J. (2005). “Probability matching priors”, in *Handbook of Statistics, 25: Bayesian Thinking: Modeling and Computation* eds. D. K. Dey and C. R. Rao, Elsevier, 91-114.

[11] Gross, A. J. and Clark, V. A. (1975), *Survival Distributions: Reliability Applications in the Biomedical Sciences* Wiley, New York.
(Section 4.8 in Chapter 4 discusses inference and model fitting for LFR as well as polynomial hazard function distributions.)

[12] Kodlin, D. (1967), “A new response time distribution”, *Biometrics*, **2**, 227–239.

[13] Krane, S. A. (1963), “Analysis of survival data by regression techniques”, *Technometrics*, **5**, 161–174.

[14] Kundu, D. (2008), “Bayesian inference and life testing plan for the Weibull distribution in presence of progressive censoring”, *Technometrics*, **50**, 144–154.

[15] Lin, C.T., Wu, S.J.S. and Balakrishnan, N. (2003), “Parameter estimation of the linear hazard rate distribution based on records and inter-record times”, *Communications in Statistics - Theory and methods*, **32**, 729 - 748.

[16] Lin, C.T., Wu, S.J.S. and Balakrishnan, N. (2006), “Monte Carlo methods for Bayesian inference on the linear hazard rate distribution”, *Communications in Statistics - Simulation and Computation*, **35**, 575 - 590.

[17] Nelson, W. (1970), “Hazard plotting methods for analysis of life data with different failure modes. *Journal of Quaity and Technology*, **2**, 126–149.

[18] Pandey, A., Singh, A. and Zimmer, W.J. (1993), “Bayes estimation of the linear failure rate model”, *IEEE Transactions on Reliability*, **42**, 636 - 640.

[19] Self, S. G. and Liang, K-Y. (1987), "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions", *Journal of the American Statistical Association*, **82**, 605–610.

[20] Sen, A. (2006), "Linear failure rate distribution", *Encyclopedia of Statistical Sciences*, 2nd. ed., John Wiley & Sons, New Jersey, 4212 – 4217.

[21] Sen, A. and Bhattacharya, G.K. (1995), "Inference procedures for linear failure rate model", *Journal of the Statistical Planning and Inference*, **46**, 59–76.

[22] Smith, A. F. M. and Gelfand, A.E. (1992), "Bayesian statistics without tears: a sampling-resampling perspective", *The American Statistician*, **46**, 84–88.

[23] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). "Bayesian measures of model complexity and fit (with discussion)", *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

[24] Viveros, R. and Balakrishnan, N. (1994), "Interval estimation of parameters of life from progressively censored data", *Technometrics*, **36**, 84–91.

[25] Zhang, Y. and Meeker, W.Q. (2005), "Bayesian life test planning for the Weibull distribution with the given shape parameter", *Metrika*, 61, 237–249.