Computations via Auxiliary Random Functions for Survival Models

Purushottam W. Laud[1] Paul Damien[2]
Stephen G. Walker[3]

[1]Division of Biostatistics, Medical College of Wisconsin, Milwaukee, USA.

[2] University of Michigan Business School, Ann Arbor, USA.

[3] Department of Mathematical Sciences, University of Bath, UK.

SUMMARY

A new simulation method, *Auxiliary Random Functions*, is introduced. When used within a Gibbs sampler, this method enables a unified treatment of exact, right-censored, left-censored, left-trucated and interval censored data, with and without covariates, in survival models. The models and methods are exemplified via illustrative analyses.

Key Words: Extended gamma process, Censored data, Gibbs sampling, Auxiliary functions, Covariates, Bioassay.

1 Introduction

The prior modelling of increasing or decreasing hazard rate functions using the extended gamma process is relatively straightforward; see, for example, Dykstra and Laud (1981). However, serious problems arise in summarising posterior distributions. Even using a Gibbs sampler, a number of computational difficulties remain (Laud et al., 1996).

In this paper, we synthesize the Dykstra and Laud (1981) model with computational ideas detailed in Damien, Wakefield and Walker (1999). These authors use auxiliary variables to simplify existing Gibbs samplers, by ensuring full conditional distributions take simple forms. Here we develop the idea of Damien et al. (1999) to solve the computational problems of Laud et al. (1996) via the use of *auxiliary random functions* (ARF). It is well known that there are unique computational difficulties associated with Bayes nonparametric models involving the Gibbs sampler; see, for example, the innovative ideas adopted by MacEachern and Müller (1998) and reviewed by MacEachern (1998) for the Dirichlet process.

2 Prior Model

Here we collect, briefly, the key properties of the extended gamma process that will be of use in the rest of the paper; see, Dykstra and Laud (1981) and Amman (1984) for further details. To begin, we consider increasing hazard rate functions. Let $G(\alpha, \beta)$ denote the gamma distribution with shape parameter $\alpha \geq 0$ and scale parameter $1/\beta > 0$. For $\alpha = 0$, we define this distribution to be degenerate at zero. For $\alpha > 0$, its density is

$$g(x|\alpha,\beta) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1}\exp(-\beta x)}{\Gamma(\alpha)}, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

Let $\alpha(t)$, $t \geq 0$, be a non-decreasing left continuous real valued function such that $\alpha(0) = 0$ and $\beta(t), t \geq 0$, be a positive right continuous real valued function with left hand limits existing and bounded away from 0 and ∞.

Take $Z(t)$, $t \geq 0$, to be a gamma process with parameter function $\alpha(\cdot)$. That is, $Z(0) = 0$, $Z(t)$ has independent increments and for $t > s$, $Z(t) - Z(s)$ is $G(\alpha(t) - \alpha(s), 1)$. Considering a version of this process such that all its sample paths are non-decreasing and left continuous, let

$$r(t) = \int_{[0,t)} [\beta(s)]^{-1} dZ(s).$$

This process $\{r(t), t \geq 0\}$ is called the Extended Gamma (EG) process and we denote it by

$$r(t) \sim \Gamma(\alpha(\cdot), \beta(\cdot)).$$

Several aspects of the posterior distribution are given by Dykstra and Laud (1981), Laud (1977) and Amman (1984). These can be summarized as follows :

Fact 2.1 : The EG process is conjugate with respect to right censored data. In particular, given an observation $X > x$, writing $\int_0^x r(t)dt$ as $\int_0^\infty (x-t)^+ dr(t)$ it follows that the posterior process is again EG with $\alpha(\cdot)$ unchanged and the β parameter modified as

$$\tilde{\beta}(t) = \beta(t) + (x - t)^+.$$

Here the superscript + on a quantity denotes its positive part.

Fact 2.2 : The posterior with respect to exact data is a mixture of EG processes. The mixing measure is n-dimensional, where n is the number of exact observations, and has both continuous and jump components.

Fact 2.3 : Although analytic expressions for the posterior mean and variance functions are given by the above mentioned authors, their numerical evaluation is virtually impossible, especially as the sample size increases to more than around 20. They do not even consider the possibility of including covariates as a consequence.

Our goal is to provide the basis for a full Bayesian analysis of the increasing (and later the decreasing and bathtub shaped) hazard rate models for various types of censored data as well as exact observations, both with and without covariates. For ease of exposition, we first focus attention on the increasing case. The rest of the discussion applies also to data arising in a bioassay because a bioassay can be treated as a censored data problem. In addition, these methods also work for estimating the intensity rate function of a nonhomogeneous Poisson process because the form of the likelihood in this case is the same as that for survival data (see, for example, Lawless (1982)).

3 Posterior Computations Via ARFs

Damien et al. (1998) develop a general method to sample from posterior distributions using auxiliary variables. Suppose the required conditional distribution for a random variable X is denoted f. The basic idea is to introduce an auxiliary variable U, construct the joint density of U and X, with marginal density for X given by f, and then extend the Gibbs sampler to include the extra full conditional for U. In a Bayesian context, consider the posterior density given by $f(x) \propto l(x)\pi(x)$ and suppose it is not possible to sample directly from f. The general idea proposed by Damien et al. is to introduce an auxiliary variable Y, defined on the interval $(0,\infty)$ or more strictly the interval $(0, l(\hat{x}))$, where $\hat{x}$ maximises $l(.)$, and define the joint density with X by

$$f(x,y) \propto I(y < l(x))\pi(x).$$

The full conditional for Y is $\mathcal{U}(0, l(x))$, a uniform variable on the interval $(0, l(x))$, and the full conditional for X is π, restricted to the set $A_y = \{x : l(x) > y\}$.

In this paper, we develop a simulation method that is particularly suited to the context of sampling the monotone hazard rate processes. Whereas Damien et al. use latent variables to put a function in place which depends on a *finite* number of parameters, the latent variables used in this paper put an entire random function into place. Hence we refer to it as the *auxiliary random function* (ARF) method. To introduce the method, consider the increasing hazard rate case without covariates, and therefore dropping the superscript I.

Exact observations. The contribution to the likelihood resulting from one exact observation at x is given by:

$$f(x) = r(x)\bar{F}(x) = r(x)\exp\left(-\int_0^x r(s)ds\right),$$

where $\bar{F}$ is the survival function. Based on Fact 2.1, it is clear that $\bar{F}(x)$, being the contribution of a right-censored observation, updates the prior parameters of the EG process. In the above equation, from an ARF perspective, the following question arises: how do we substitute the random function $r(\cdot)$ so that sampling from the posterior is accomplished readily? Take the cumulative distribution of a random variable Y given the hazard rate process $r(\cdot)$ as:

$$P(Y \leq y|r(\cdot)) = \frac{r(y)}{r(x)} I(0 \leq y \leq x) + I(y \geq x).$$

The distribution of Y is defined in terms of a random function, hence the phrase "auxiliary random function". It is in this sense that the ARF algorithm is different from Damien et al.'s algorithm. To sample $[Y|r(\cdot)]$, let $Y = r^{-1}(Ur(x))$ where $U \sim \mathcal{U}(0,1)$ and $r^{-1}(w) = inf\{t|r(t) \geq w\}$. To address sampling $[r(\cdot)|Y = y]$, write the density of $Y|r(\cdot)$ as

$$[y|r(\cdot)] = \frac{dr(y)}{r(x)} I(0 \leq y \leq x).$$

Combining this with the prior and the likelihood we arrive at

$$[r(\cdot)|y] \propto [r(\cdot)] dr(y) \exp\left(-\int_0^x r(s)ds\right),$$

where $dr(y)$ denotes the increment of $r(\cdot)$ at y. It is clear that this term combines naturally with the EG process, adding a unit jump at y to the shape parameter $\alpha(\cdot)$. Thus, as a result of the ARF substitution and Fact 2.1, the posterior full conditional for $r(\cdot)$ is once again an EG process with the parameters updated by:

$$\tilde{\alpha}(s) = \alpha(s) + I(s \geq y) \text{ and } \tilde{\beta}(s) = \beta(s) + (x-s)^+.$$

The full conditional distribution of the hazard rate function after the ARF substitution reduces to an EG process. The increments of the EG process can be simulated by using the Bondesson (1982) method as described in Laud, Smith and Damien (1996) or, as is often adequate, simply approximated as gamma variates.

Left censored data. The contribution to the likelihood from a left-censored observaton, $X \leq x$, is

$$F(x) = 1 - \exp\left(-\int_0^x r(s)ds\right).$$

First, using the auxiliary variate $[W|r(\cdot)]$ with hazard rate $r(\cdot)$ restricted to $[0, x]$, we have

$$[w|r(\cdot), X \leq x] = \frac{r(w)\exp(-\int_0^w r(s)ds)}{1 - \exp(-\int_0^x r(s)ds)} I(0 \leq w \leq x),$$

and

$$[r(\cdot)|w, X \leq x] \propto [r(\cdot)] r(w) \exp\left(-\int_0^w r(s)ds\right).$$

Now, in a manner similar to the exact data scenario, we introduce an ARF Y and proceed as described above.

Interval censored data. The contribution of an observation, $a < X \leq b$ to the likelihood is

$$\bar{F}(a) - \bar{F}(b) = \exp\left(-\int_0^a r(s)ds\right) - \exp\left(-\int_0^b r(s)ds\right).$$

Employing $W|r(\cdot)$ with hazard rate $r(\cdot)$ restricted to $(a,b]$ yields the same full conditional for $r(\cdot)$ as in the left censored data scenario above.

Left truncated data. Here one observes X (or a censored version of it) only if it exceeds an observed left truncation time L. Thus the contribution of $X = x, L = l$ to the likelihood is

$$\frac{f(x)}{\bar{F}(l)} I(l \leq x) = r(x)\exp\left(-\int_l^x r(t)dt\right) I(l \leq x).$$

Once again, the ARF Y as defined above suffices to substitute for $r(x)$. It can be shown easily that the effect of the term $e^{-\int_l^x r(t)dt}$ is to modify the scale functions given in Fact 2.1 as

$$\tilde{\beta}(t) = \beta(t) + (x - max(t,l))^+.$$

4 Cox Model Computations Via ARFs

Consider now the extension of the nonparametric analysis to include covariates. Employing the Cox model, Cox (1972), and initially tackling a single right-censored observation, let β and z denote, respectively, the vectors of regression parameters and covariates. The likelihood function for a right censored observation is given by

$$L(r(\cdot),\beta) = P(X > x|r(\cdot),\beta) = \exp\left(-e^{z'\beta}\int_0^x r(s)ds\right).$$

Conditioning on β and following the development leading to Fact 2.1, we easily arrive at

$$[r(\cdot)|\beta, X > x] \sim EG\left(\alpha(\cdot), \tilde{\beta}(\cdot)\right)$$

where

$$\tilde{\beta}(\cdot) = \beta(\cdot) + e^{z'\beta}(x - s)^+.$$

In the case of exact data,

$$[r(\cdot)|\beta, X = x] = [r(\cdot)]e^{z'\beta}r(x)\exp\left(-e^{z'\beta}\int_0^x r(s)ds\right).$$

Clearly, the problem now yields to the same ARF substitution employed in the case without covariates. To examine the full conditional distribution of β, let δ denote the usual "noncensoring" indicator and write the likelihood as

$$L(r(\cdot),\beta; x,\delta) \propto e^{z'\beta\delta}\exp\left(-e^{z'\beta}\int_0^x r(s)ds\right).$$

Simulating from $[\beta|\cdot]$ is easily accomplished using the adaptive rejection sampling algorithm of Gilks and Wild (1992) based on the log-concavity of $[\beta|\cdot]$, having assigned independent normal priors for the β components.

Sampling $r(\cdot)$, which avoids the grid approximation methods of Bondesson (1982) and Damien et al. (1995), is available using representations of Lévy processes detailed in, for example, Walker and Damien (2000).

5 Illustrative Analysis

For illustrative purposes we present here an example using a simulated data set of size $n = 500$. We take the simulation model to be a Weibull proportional hazards model; that is,

$$h_i(t) = h_0(t) \exp(\beta_1 z_{i1} + \beta_2 z_{i2}),$$

where $h_0(t) = t^2/7200$, $\beta_1 = 0.405$ and $\beta_2 = 0.693$. The z_{ij}s are taken to be in $\{0, +1\}$ independently generated and $P(z_{ij} = 0) = 0.5$.

We took independent normal priors for β_1 and β_2 with zero means and standard deviation of 10, resulting in a non-informative set-up. The prior parameter functions $\alpha(\cdot)$ and $\beta(\cdot)$ were chosen so that $r(\cdot)$ was centred on a Weibull hazard rate function but with a large variance.

The Gibbs sampler was run for 4000 iterations with a burn-in of 2000 leaving 2000 samples from the posterior distributions with which to make posterior summaries.

In the Figure we present the estimate of the baseline hazard rate function, with associated 2 standard deviation limits alongside the true hazard rate function. It is evident that the algorithm has performed well in this case.

6 Decreasing and Bathtub shaped hazard functions

In a manner similar to the description of the increasing hazard rate fucntion, one can define an EG process for decreasing hazard rates as in Laud (1977). Take $\alpha(\cdot)$, $\beta(\cdot)$ and $Z(\cdot)$ as with the increasing case. In addition, let $Z(\infty) \sim G(\alpha(\infty), 1)$ be such that $Z(\infty) - \lim_{t\to\infty} Z(t)$ is independent of the rest of the process, $\alpha(\infty) > \lim_{t\to\infty} \alpha(t)$ and $\beta(\infty) > 0$. Now define

$$r(t) = \int_{[t,\infty]} [\beta(s)]^{-1} dZ(s) = \int_{[t,\infty)} [\beta(s)]^{-1} dZ(s) + [\beta(\infty)]^{-1} [Z(\infty) - \lim_{t\to\infty} Z(t)] \quad .$$

The main purpose of the values at infinity is to preclude $\lim_{t\to\infty} r(t) = 0$ so that $\int_{[0,\infty)} r(t)dt$ is infinite and the survival time described by $r(\cdot)$ does not have positive probability at infinity. It is clear that such a process generates non-increasing hazard rates. We call it the Decreasing Extended Gamma (DEG) and denote it by

$$r(t) \sim D\Gamma(\alpha(\cdot), \beta(\cdot)).$$

Amman (1984) uses the above development and defines a *combined* process by

$$r(t) = r^I(t) + r^D(t),$$

where the superscript, inherited also by the parameter functions, indicates the direction of monotonicity, and the two component processes are independent. Amman termed this a U-shaped process whereas elsewhere in the reliability and survival analysis literature one also encounters the usage "bathtub-shaped hazard rates". It should be noted that such a combined process does not necessarily generate hazard rates with this shape. It does, however, add a great deal of flexibility in modelling the shape. For example, the expected hazard rate can be arranged to decrease initially and to increase beyond a certain time point by a suitable choice of the parameter functions. A parametric choice for the expectation can be a Weibull distribution with shape parameter less than one for the decreasing component and shape parameter exceeding two for the increasing component. In general, one can require

$$\frac{d^2}{dt^2}E(r(t)) = \frac{d^2}{dt^2}E(r^I(t)) + \frac{d^2}{dt^2}E(r^D(t)) \geq 0$$

with

$$\frac{d}{dt}E(r(t)) = 0 \text{ for some } t > 0.$$

Since

$$E(r(t)) = \int_{[0,t)} [\beta^I(s)]^{-1} d\alpha^I(s) + \int_{[t,\infty]} [\beta^D(s)]^{-1} d\alpha^D(s),$$

we get the conditions

$$\frac{\beta^I(t)\alpha^I_{(2)}(t) - \alpha^I_{(1)}(t)\beta^I_{(1)}(t)}{[\beta^I(t)]^2} - \frac{\beta^D(t)\alpha^D_{(2)}(t) - \alpha^D_{(1)}(t)\beta^D_{(1)}(t)}{[\beta^D(t)]^2} \geq 0$$

and

$$\frac{\alpha^I_{(1)}(t)}{\beta^I(t)} - \frac{\alpha^D_{(1)}(t)}{\beta^D(t)} = 0 \text{ for some } t > 0$$

where the parenthesized subscript i on a function denotes its i^{th} derivative. Such a prior, while maintaining a bathtub expectation, allows the hazard rates to have more general shapes. The data can then either reinforce the bathtub shape or indicate otherwise via the posterior expectation.

Facts 2.1 to 2.3 also apply to the decreasing hazard rate function, except for $\tilde{\beta}^D(t) = \beta^D(t) + min(x,t)$ in the decreasing hazard case and for Fact 2.1.

The case of the decreasing hazard rates is similar : the auxiliary variable Y has the survival function

$$P(Y > y|r(\cdot)) = \frac{r(y)}{r(x)} I(y > x)$$

and possibly takes the value "∞". This results in a DEG process for the full conditional for $r(\cdot)$ with parameters

$$\tilde{\alpha}(s) = \alpha(s) + I(s \geq y) \text{ and } \tilde{\beta}(s) = \beta(s) + min(x, s).$$

Turning to the case of the combined process, we write the posterior distribution as

$$[r^D(\cdot), r^I(\cdot)|X = x] \propto [r^D(\cdot)][r^I(\cdot)]\{r^D(x) + r^I(x)\} \exp(-\int_0^x \{r^D(s) + r^I(s)\}ds).$$

Now the stochastic substitution takes the form

$$[y|X = x, r^D(\cdot), r^I(\cdot)] = \frac{dr^I(y)I(0 \leq y \leq x) + \{-dr^D(y)\}I(y > x)}{r^D(x) + r^I(x)}.$$

Straightforward algebra results in

$$[r^D(\cdot), r^I(\cdot)|y, X = x] \propto EG(\tilde{\alpha}^I, \tilde{\beta}^I)EG(\tilde{\alpha}^D, \tilde{\beta}^D),$$

where

$$\tilde{\alpha}^D(s) = \alpha^D(s) + I(y > x)I(s \geq y), \quad \tilde{\alpha}^I(s) = \alpha^I(s) + I(0 \leq y \leq x)I(s \geq y)$$

and the two scale parameter functions as given above.

The updates for left censored data, interval censored data and left truncated data follow straighforwardly.

7 Discussion

In this paper we have provided a complete and easy to implement Bayesian solution to modelling monotone increasing hazard rate functions nonparametrically. We point out that the approach readily extends to the modelling of monotone decreasing and bathtub hazard rate functions based on the EG process (see, Laud, 1977, and Amman, 1984). A new stochastic simulation method was introduced to sample the posterior process. We also provided a solution for the semi-parametric model with covariates. Natural extensions are to consider time-dependent covariates and to detect when data contradict the proportional hazards assumption. From the perspective of the physicians and scientists wishing to use these methods, based on the stored posterior samples, graphical implementations that can respond to queries regarding substantively interesting functionals would be attractive indeed.

References

Amman, L. (1984). Bayesian nonparametric inference for quantal response data. *Annals of Statistics*, **12**, 636–645.

Bondesson, L. (1982). On simulation from infinitely divisible distributions. *Adv. App. Prob.* 14, 855–869.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B*, 34, 187–220.

Damien, P., Wakefield, J.C. and Walker, S.G. (1998). Gibbs sampling for Bayesian nonconjugate models using auxiliary variables. *J. Roy. Statist. Soc. B* **61**, 331–344.

Dykstra, R.L. and Laud, P.W. (1981). A Bayesian nonparametric approach to reliability. *Annals of Statistics*, **9**, 356–367.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

Laud, P.W. (1977). Bayesian nonparametric inference in reliability. Ph.D. Dissertation, University of Missouri, Columbia, MO.

Laud, P.W., Smith, A.F.M. and Damien, P. (1996). Monte Carlo methods for approximating a posterior hazard process. *Statistics and Computing*, **6**, 77–83.

Lawless, J.F. (1982). *Statistical Models and Methods For Lifetime Data*, Wiley Publications, New York.

MacEachern, S.N. and Müller, P. (1998). Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238.

MacEachern, S.N. (1998). Computational methods for mixtures of Dirichlet process models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (Eds. D.Dey, P.Müller and D.Sinha.) Chapter 2, Springer.

Walker, S.G. and Damien, P. (2000). Representations of Lévy processes without Gaussian components. *Biometrika* **87**, 477–483.

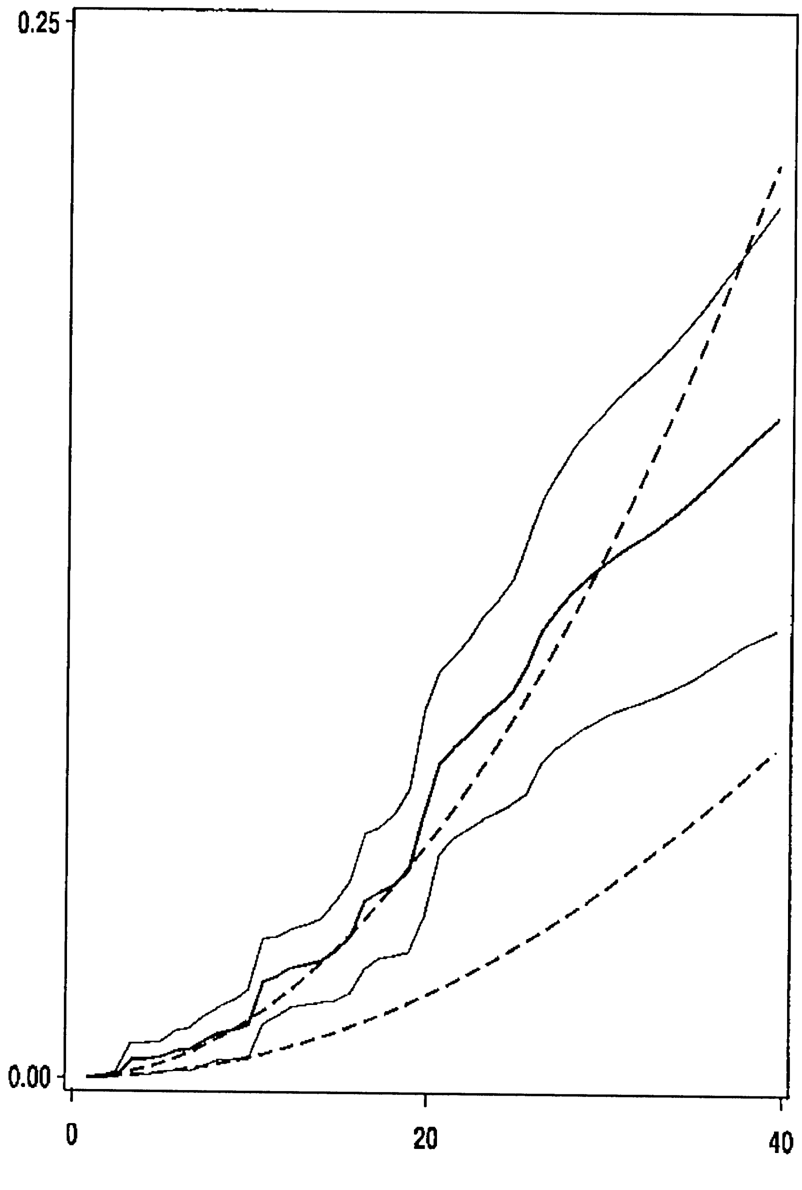

Figure. Solid lines represent posterior mean and two standard deviation limits. Lower dashed line shows the prior mean, the upper the true hazard.