

EXTENDED HOMOGENEOUS PROCESSES AND BAYES ESTIMATION OF RELIABILITY FUNCTIONS

L. PARDO, D. MORALES AND V. QUESADA

Abstract. The problem of estimation a reliability function is established in the Bayesian nonparametric context; however parametric techniques are used. Extended homogeneous processes are defined whose sample paths may be assumed to be increasing hazard rates by properly choosing the parameter functions of the processes. Estimators are obtained in the mentioned processes and their asymptotic properties are studied. An application for simulated data is given.

1. Introduction

Suppose the random variable $T \geq 0$ denotes the age of failure of an inanimate device. The reliability function and the cumulative hazard function corresponding to T are defined by

$$R(t) = Pr(T > t)$$

and

$$H(t) = -\log R(t)$$

respectively. If for all $t \geq 0$

$$H(t) = \int_{[0,t)} r(u) du$$

then $r(t)$ is called the hazard rate of the distribution and in reliability context indicates the propensity for failure of an item in the near future given that the item has survived till time t .

We use a Bayesian nonparametric approach to estimate a reliability function; this is, we do not introduce significant constraints over the space \mathcal{R} of all the reliability functions.

A problem, which appears immediately, is to place a prior probability over \mathcal{R} . T. S. Ferguson (1973) and K. Doksum (1974) have studied this problem by defining the

Received November 21, 1990.

Key words and phrases: Hazard rates, increasing hazard rates, Bayes estimates, extended homogeneous processes, prior processes.

AMS 1980 subject classifications: Primary 62N05, secondary 62F51.

Dirichlet process and the more general neutral to the right processes respectively. An alternative solution is given by H. L. Dikstrat et al. (1981). They have introduced an appropriate stochastic process, called extended gamma process, whose sample paths are hazard rates. Moreover, they derive its posterior distribution and give the Bayes estimator or $r(t)$ under the natural loss function

$$L(r, \hat{r}) = \int_0^\infty [r(t) - \hat{r}(t)]^2 dW(t)$$

where W is an arbitrary finite measure on $[0, \infty)$.

In accordance with these last ideas, we use extended homogeneous processes, which have the advantage of placing the prior probability on absolutely continuous rather than on discrete distributions, as is the case with a neutral to the right process prior not having a nonrandom part.

In this work, we present a Bayesian parametric method to estimate $R(t)$, where it is not necessary to calculate the posterior distribution of the process. This method is similar to one presented in Morales et al. (1986).

2. The Extended Homogeneous Processes

Let $Z(t)$, $t \geq 0$, defined on an appropriate probability space (Ω, A, P) , denote a process with independent increments such that:

- i) $Z(0) = 0$ a.s.
- ii) $Z(t)$ has nondecreasing right continuous sample paths a.s.
- iii) $\lim_{t \rightarrow \infty} Z(t) = +\infty$ a.s.
- iv) The characteristic function of $Z(t)$ has the following Levy-Kolmogorov representation:

$$\log \psi_t(\theta) = \tau(t) \int_0^{+\infty} [e^{i\theta z} - 1] dN(z)$$

where $\tau(t)$, $t \geq 0$, is a nondecreasing right-continuous real-valued function such that $\tau(0) = 0$, and $N(z)$ is a an arbitrary measure on $[0, \infty)$ such that

$$\int_0^{+\infty} \frac{z}{1+z} dN(z) < \infty \quad \text{and} \quad \int_0^{+\infty} z^2 dN(z) < \infty$$

Let $\beta(t)$, $t \geq 0$, be a positive right-continuous real valued functions, bounded away from 0 with left-hand limit existing. We denote a new stochastic process by

$$r(t) = \int_{[0,t)} \beta(s) dZ(s)$$

where the integration is with respect to the sample paths of the $Z(t)$ process. We say a process defined in this manner has an extended homogeneous distribution. Of course if

$r(t)$ is taken to be a random hazard rate, there will be a corresponding random continuous reliability function given by

$$R(t) = \exp\left[-\int_{[0,t)} r(u) du\right]$$

and we denote such a process by saying $r(t)$ or $R(t)$ is $H(\tau(t), \beta(t), N(z))$. Furthermore, it can be easily proved that

$$E[r(t)] = k_1 \int_{[0,t)} \beta(s) d\tau(s)$$

and

$$V[r(t)] = k_2 \int_{[0,t)} \beta^2(s) d\tau(s)$$

where

$$k_1 = \int_0^{+\infty} z dN(z) \quad \text{and} \quad k_2 = \int_0^{+\infty} z^2 dN(z) < \infty$$

In this work, we consider two special extended homogeneous random reliability functions:

a) Extended gamma process

$$dN(z) = dz/(ze^z) \quad \text{and} \quad k_1 = k_2 = 1$$

b) Extended simple process

$$dN(z) = dz/(e^z) \quad \text{and} \quad k_1 = 1, k_2 = 2$$

In assigning a prior probability measure to \mathcal{R} , one needs to input the functions $\tau(t)$ and $\beta(t)$. One approach consists of defining nondecreasing mean and variance functions $\mu(t)$ and $\sigma^2(t)$. It would seem reasonable to assign as $\mu(t)$ the best "guess" of the hazard rate and use $\sigma^2(t)$ to measure the amount of uncertainty or variation in the hazard rate at the point t . Assuming $\mu(t)$, $\sigma^2(t)$ and $\beta(t)$ are all differentiable, one can set $\mu(t) = E[r(t)]$ and $\sigma^2(t) = V[r(t)]$.

Solving for $\tau(t)$ and $\beta(t)$ yields,

$$\beta(t) = \frac{k_1 (\sigma^2(t))'}{k_2 \mu'(t)} \quad \text{and} \quad \tau'(t) = \frac{k_2 (\mu'(t))^2}{(k_1)^2 (\sigma^2(t))} \tag{2.1}$$

which then determines the prior distribution.

3. A Parametric Method to Estimate a Reliability Function

For each fixed $t \geq 0$, we consider the following parametric Bayesian estimation problem. The parameter $R(t)$ is a random variable whose p.d.f. is given by the extended homogeneous process $\{R(s)\}_{s \geq 0}$. This p.d.f. will be the prior probability for this particular problem. We consider, for the mentioned $t \geq 0$, the following random variable:

$$X^t = \begin{cases} 1 & \text{if } T > t \\ 0 & \text{if } T \leq t \end{cases}$$

After observing a random sample T_1, \dots, T_n from $T \geq 0$, we have the associated sample X_1^t, \dots, X_n^t from X^t . So, $nR_n(t) = X_1^t + \dots + X_n^t$ has a Binomial distribution with parameters $(n, R(t))$.

If we use the quadratic loss function $L[R(t), \hat{R}_n(t)] = [R(t) - \hat{R}_n(t)]^2$ the Bayes estimate of $R(t)$, for the fixed t , which minimizes the expected loss is given by the posterior mean of $R(t)$.

Observe that we do not need to calculate, with this method, the posterior distribution of the random reliability function.

Theorem 3.1 *If $\{r(s)\}_{s \geq 0}$ is distributed as $H(\tau(t), \beta(t), N(t))$ then, for each fixed $t \geq 0$, the Bayes estimator of $R(t)$ under quadratic loss is given by the expression*

$$\hat{R}_n(t) = \frac{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \exp\left(\int_{[0,t)} \int_0^{+\infty} (e^{-\beta(s)(t-s)} z^{(n-j+1)} - 1) dN(z) d\tau(s)\right)}{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \exp\left(\int_{[0,t)} \int_0^{+\infty} (e^{-\beta(s)(t-s)} z^{(n-j)} - 1) dN(z) d\tau(s)\right)}$$

where x denotes $nR_n(t)$.

Proof: For each fixed $t \geq 0$, let $g(y)$ be the p.d.f. of the random variable $H(t)$. If we denote $R(t)$ by θ , $p_t(\theta)$ is the p.d.f. of $R(t)$ and

$$f(x_1, x_2, \dots, x_n / \theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$$

is the joint density of the n random variables X_1^t, \dots, X_n^t . Finally, we define $x = \sum_{i=1}^n x_i$.

Using the binomial expansion, the posterior mean of $R(t)$ is

$$\begin{aligned} \hat{R}_n(t) = E[\Theta / nR_n(t)] &= \frac{\int_0^1 \theta^{x+1} (1 - \theta)^{n-x} dp_t(\theta)}{\int_0^1 \theta^x (1 - \theta)^{n-x} dp_t(\theta)} \\ &= \frac{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \int_0^{+\infty} e^{-(n-j+1)y} g(y) dy}{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \int_0^{+\infty} e^{-(n-j)y} g(y) dy} \end{aligned}$$

Observe that

$$\int_{[0,\infty)} e^{-ay} g(y) dy = E[\exp(-aH(t))] = E[\exp(-a \int_{[0,\infty)} r(s) ds)].$$

Now, this expectation can be calculated by defining a sequence of partitions $0 = t_{n,1} < t_{n,2} < \dots < t_{n,k(n)}$ whose norm goes to 0 and whose upper end point goes to ∞ . Then the sequence of random functions

$$r_n(t) \begin{cases} \sum_{t_{n,1} < t} \beta(t_{n,i}) (Z(t_{n,i}) - Z(t_{n,i-1})) & \text{if } t > t_{n,2} \\ 0 & \text{if } t_{n,1} < t < t_{n,2} \end{cases}$$

verifies that $\lim_{n \rightarrow \infty} r_n(t) = r(t)$ a.s.

Appropriate limiting arguments gives the following result:

$$E[\exp(-aH(t))] = \exp\left[\int_0^\infty \left(\int_0^t (e^{-\beta(s)z(t-s)(n-j+1)} - 1) d\tau(s) dN(z)\right)\right]$$

This completes the proof.

If we particularize the last theorem for some specific extended homogeneous processes, we obtain:

-Extended Gamma

$$\hat{R}_n(t) = \frac{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \exp\left[-\int_{[0,t)} \log(1 + \beta(s)(t-s)(n+1-j)) d\tau(s)\right]}{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \exp\left[-\int_{[0,t)} \log(1 + \beta(s)(t-s)(n-j)) d\tau(s)\right]}$$

-Extended Simple

$$\hat{R}_n(t) = \frac{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \exp\left[-\int_{[0,t)} \frac{\beta(s)(t-s)(n-j+1)}{\beta(s)(t-s)(n-j+1)+1} d\tau(s)\right]}{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \exp\left[-\int_{[0,t)} \frac{\beta(s)(t-s)(n-j)}{\beta(s)(t-s)(n-j)+1} d\tau(s)\right]}$$

4. Some Properties of $\hat{R}_n(t)$

Theorem 3.1 gives, for each fixed $t \geq 0$, a Bayes estimator of $R(t)$. Now the problem is to consider the function $\hat{R}_n(\cdot)$ as an estimator of the function $R(\cdot)$. To do this, $\hat{R}_n(\cdot)$ must be a deterministic reliability function. The next theorem states this result.

Theorem 4.1. $\hat{R}_n(\cdot)$ is a deterministic reliability function.

Proof: First we prove that $\lim_{t \rightarrow 0} \hat{R}_n(t) = 1$ and $\lim_{t \rightarrow \infty} \hat{R}_n(t) = 0$ a.s.. If t goes to 0 or ∞ then $r(t)$ goes to 1 or 0 a.s., so the prior probability $p_t(\theta)$ will concentrate its mass in those points. As the posterior probability $p_t(\theta/x)$ is proportional to the prior, we obtain that $p_t(\theta/x)$ will concentrate its mass in the mentioned points as well. Considering that $R(\cdot)$ is bound a.s, the mean of $p_t(\theta/x)$ will be 1 or 0 respectively.

Second we prove that $\hat{R}_n(\cdot)$ is a right-continuous function. $\hat{R}_n(\cdot)$ is a quotient of linear combinations of moment-generating functions which are continuous. The denominator is never 0, the jumps of $\hat{R}_n(\cdot)$ are at the sample values t_1, \dots, t_n , and $\hat{R}_n(\cdot)$ is continuous in $[t_i, t_{i-1})$.

Finally we prove that $\hat{R}_n(\cdot)$ is a nondecreasing function.

$\hat{R}_n(t) = E[R(t)/nR_n(t)]$ is a $\sigma(nR_n(t))$ -measurable function and $nR_n(t)$ is a discrete random variable taking the values $0, 1, \dots, n$. Therefore $\hat{R}_n(t)$ is a constant in each set $A_x = \{\omega \in \Omega/nR_n(t)(\omega) = x\}$. This is,

$$\forall \omega \in A_x \quad R_n(t)(\omega) = r_x = \int_0^1 \theta dp_t(\theta/x), \quad x = 0, 1, \dots, n$$

First we prove that $r_1 < r_2 < \dots < r_n$. We study the behavior of the posterior distribution function of $R(t)$ given $nR_n(t) = x$.

$$\begin{aligned} dp_t(\theta/(x+1)) > dp_t(\theta/x) &\Leftrightarrow \frac{\binom{n}{x+1} \theta^{x+1} (1-\theta)^{n-x-1} dp_t(\theta)}{\int_0^1 \binom{n}{x+1} \theta^{x+1} (1-\theta)^{n-x-1} dp_t(\theta)} \\ &> \frac{\binom{n}{x} \theta^x (1-\theta)^{n-x} dp_t(\theta)}{\int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} dp_t(\theta)} \Leftrightarrow \frac{\int_0^1 \theta^x (1-\theta)^{n-x} dp_t(\theta)}{\int_0^1 \theta^{x+1} (1-\theta)^{n-x-1} dp_t(\theta)} \\ &> \frac{\theta^x (1-\theta)^{n-x} dp_t(\theta)}{\theta^{x+1} (1-\theta)^{n-x-1} dp_t(\theta)} = \frac{1-\theta}{\theta} \end{aligned}$$

If and only if $\exists b \in (0, 1)$ s.t $\theta > b$, because $\frac{1-\theta}{\theta}$ is decreasing in $(0, 1)$, satisfying $\lim_{\theta \rightarrow 1} \frac{1-\theta}{\theta} = 0$. Therefore, $r_{x+1} = E_{p_t(\theta/(x+1))}(\theta) \geq E_{p_t(\theta/x)}(\theta) = r_x \quad \forall x = 0, 1, \dots, n-1$. Now we compare $\hat{R}_n(t)$ with $\hat{R}_n(t')$ when $t > t'$. We know that $\hat{R}_n(t')$ is a constant in the set $A'_x = \{\omega \in \Omega/nR_n(t')(\omega) = x\}$ this is, $\forall \omega \in A'_x \quad \hat{R}_n(t')(\omega) = r'_x = \int_0^1 \theta dp_t(\theta/x)$.

Furthermore, as $nR_n(t') \leq nR_t(t)$, the sets A'_x satisfy $A'_x \subset \bigcup_{i=x}^n A_i$; so it is enough to prove the following inequalities $r'_x < r_i \quad \forall x = 1, \dots, i$ However, $r'_0 < r'_1 < \dots < r'_n$, it

is only necessary to prove that $r'_0 < r_0, r'_1 < r_1, \dots, r'_n < r_n$. We analyze the relation between the posterior p.d.f. $p_t(\theta/x)$ and $p_{t'}(\theta/x)$.

$$dp_t(\theta/x) > dp_{t'}(\theta/x) \Leftrightarrow \frac{dp_t(\theta)}{dp_{t'}(\theta)} > \frac{P(nR_n(t) = x)}{P(nR_n(t') = x)} = k \in (0, \infty)$$

We observe that $R(t) > R(t')$ a.s and suppose, to simplify reasoning, that $p_t(\theta)$ and $p_{t'}(\theta)$ are unimodal continuous p.d.f..

In the case, as $\theta^x(1 - \theta)^{n-x}$ is a concave function in $(0, 1)$ with maximum in x/n , the posterior p.d.f. $p_t(\theta/x)$ and $p_{t'}(\theta/x)$ have the same structure.

If consider the set $B_k = \{\theta \in [0, 1] / dp_t(\theta) > kdp_{t'}(\theta)\}$, B_k is a nonempty set because $dp_{t'}(\theta/x)$ cannot be greater than $dp_t(\theta/x)$ for every $\theta \in (0, 1)$.

We obtain $B_k = (b, 1)$, with $b \geq \theta_0$ when $k \geq 1$ and $b < \theta_0$ when $k < 1$. Therefore, $r'_x = E_{p_{t'}(\theta/x)}(\theta) < E_{p_t(\theta/x)}(\theta) = r_x$.

Now we give a result related to the Bayesian behavior of $\hat{R}_n(t)$. Remember that to determine the parameter functions $\tau(t)$ and $\beta(t)$ of an extended homogeneous process it is necessary to give two functions: $\mu(t)$ and $\sigma^2(t)$. $\mu(t)$ represents the prior knowledge about $r(t)$ and $\sigma^2(t)$ can be considered as a measure of how confident we are about that knowledge. If $\mu(t)$ remains constant and $\sigma^2(t)$ goes to 0, which means that our trust grows, it seems reasonable that $\hat{R}_n(t)$ goes to the reliability function associated to the prior estimation of $r(t)$; this is,

$$R_0(t) = \exp\left(-\int_0^t \mu(s) ds\right)$$

Theorem 4.2 Suppose $\mu(0) = 0$. For every fixed $t \geq 0$, if $\mu(t)$ remains constant, then

$$\lim_{\sigma^2(t) \rightarrow 0} \hat{R}_n(t) = R_0(t)$$

Proof: As,

$$\mu(t) = k_1 \int_0^t \beta(s)\tau'(s) ds$$

remains constant and

$$\sigma^2(t) = k_2 \int_0^t \beta^2(s)\tau'(s) ds$$

goes to 0, we have that $\beta(s)\tau'(s)$ remains constant for every $s \leq t$ and $\beta(s)$ goes to 0. Therefore, using standard calculus techniques, we obtain

$$\lim_{\beta(s) \rightarrow 0} \hat{R}_n(t) = R_0(t)$$

5. Computation and Simulation. An Example

In this section random samples of size $n = 10$ from an exponential distribution with parameter $a = 1$, i.e. $R(t) = e^{-t}$, are generated in order to calculate $\hat{R}_n(t)$. We suppose that the prior estimation of $R(t)$ is $R_0(t) = e^{-t^2}$, i.e., Weibull with parameters $a = 1$ and $b = 2$.

As it was stated in section 2, in assigning a prior probability measure to \mathcal{R} , one needs to input the mean and variance functions of $r(t)$, i.e., $\mu(t)$ and $\sigma^2(t)$ respectively. Remember that $\mu(t) = E[r(t)] = r_0(t) = 2t$ is the best guess of the hazard rate function and $\sigma^2(t)$ indicates how confident we are in such a guess. Furthermore, we suppose that $\sigma^2(t) = 4t$. We consider two different prior processes: (1) Extended Gamma (2) Extended Simple.

If the prior process is Extended Gamma, solving equations (2.1) for $k_1 = k_2 = 1$, $\mu(t) = 2t$ and $\sigma^2(t) = 4t$, we obtain $\beta(t) = 2$ and $\tau(t) = t$, so

$$\hat{R}_n(t) = \frac{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} [1 + (n+1-j)2t]^{-\left(t + \frac{1}{2(n+1-j)}\right)}}{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} [1 + (n-j)2t]^{-\left(t + \frac{1}{2(n-j)}\right)}}$$

Computational results for this process are summarized in table 1.

If the prior process is Extended Simple, solving equations (2.1) for $k_1 = 1$, $k_2 = 2$, $\mu(t) = 2t$ and $\sigma^2(t) = 4t$, we obtain $\beta(t) = 1$ and $\tau(t) = 2t$, so

$$\hat{R}_n(t) = \frac{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} [1 + (n+1-j)t]^{\frac{2}{n+1-j}}}{\sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} [1 + (n-j)t]^{\frac{2}{n-j}}}$$

Computational results for this process are summarized in table 2.

Finally, in both tables t indicates the ordered simulated values from $R(t)$, $\hat{R}_{10}(t)$ is the estimation of $R(t)$, $\hat{R}_{10}(t^-) = \lim_{s \uparrow t} \hat{R}_{10}(s)$, $R(t) = e^{-t}$ and $R_0(t) = e^{-t^2}$.

Table 1.

t	$\hat{R}_{10}(t)$	$\hat{R}_{10}(t^-)$	$R_0(t)$	$R(t)$
0.08214	0.95009	0.99673	0.99327	0.92114
0.14735	0.87028	0.93411	0.97882	0.86299
0.38806	0.74492	0.82353	0.86020	0.67837
0.48242	0.64970	0.73122	0.79237	0.61729
0.51346	0.56234	0.64528	0.76825	0.59842
0.53135	0.47562	0.55972	0.75402	0.58781
0.64331	0.37438	0.46012	0.66110	0.52555
0.70336	0.27989	0.36691	0.60974	0.49492
0.80295	0.18014	0.26870	0.52480	0.44801
3.35460	0.05938	0.06033	0.00001	0.03492

Table 2.

t	$\hat{R}_{10}(t)$	$\hat{R}_{10}(t^-)$	$R_0(t)$	$R(t)$
0.08214	0.93590	0.99712	0.99327	0.92114
0.14735	0.85770	0.91230	0.97882	0.86299
0.38806	0.72864	0.79840	0.86020	0.67837
0.48242	0.63080	0.70840	0.79237	0.61729
0.51346	0.54643	0.62526	0.76825	0.59842
0.53135	0.48415	0.54328	0.75402	0.58781
0.64331	0.36408	0.44515	0.66110	0.52555
0.70336	0.27319	0.35485	0.60974	0.49492
0.80295	0.17692	0.25925	0.52480	0.44801
3.35460	0.04421	0.07308	0.00001	0.03492

References

- [1] Dykstra, H. L. and Laud, P., "A Bayesian nonparametric approach to reliability", *Ann. Statist.* 9, n. 2, 356-367, 1981.
- [2] Doksum, K., "Tailfree and neutral random probabilities and their posterior distributions", *Ann. Probability* 2, 183-201, 1974.
- [3] Ferguson, T. S., "A Bayesian analysis of some nonparametric problems", *Ann. Statist.* 1, 209-230, 1973.
- [4] Morales, D., Quesada, V. and Pardo, L., "Estimación paramétrica bayesiana no paramétrica de funciones de supervivencia con observaciones parcialmente censuradas", *Trabajos de Estadística* 1, n. 1, 70-87, 1986.
- [5] Morales, D., Pardo, L. and Quesada, V., "Estimation of a survival function with doubly censored data and Dirichlet process prior knowledge on the observable variable", *Comm. in Statist. (Sim. and comp.)* 1, 349-362, 1990.

Departamento de Estadística e I.O., Facultad de Matemáticas.
 Universidad Complutense de Madrid, 28040-Madrid, Spain.