⊠연구논문

고장률 함수의 평활추정*

나명환 · 이현우 · 김재주
서울대학교 통계학과

A Smooth Estimation of Failure Rate Function

Myoung-Hwan Na · Hyun-woo Lee · Jae-Joo Kim
Dept. of Statistics Seoul National University

Abstract

We introduce a method of estimating an unknown failure rate function based on sample data. We estimate failure rate function by a function s from a space of cubic splines constrained to be linear (or constant) in tails using maximum likelihood estimation. The number of knots are determined by Bayesian Information Criterion(BIC). Examples using simulated data are used to illustrate the performance of this method.

1. Introduction

Reliability engineers, biostatisticians, and actuaries are all interested in lifetimes. In particular, they are interested in five lifetime distribution representations: the failure rate function $\lambda(t)$, the cumulative failure rate function $\Lambda(t)$, the reliability function $R(t)$, the probability density function $f(t)$, and the mean residual life function $m(t)$.

Perhaps, the failure rate function is the most popular of the five representations for life-time modelling. The first reason for its popularity is its intuitive interpretation as the amount of risk associated with an item at time t. The second

* 본 연구는 1997년도 교육부 학술 연구 조성비 BSRI-97-1415에 의하여 연구되었음.

reason is its usefulness in comparing the way risks change over time for several populations of items by plotting their failure rate functions on a single axis. The third reason is that the failure rate function is a special case of the intensity function for a nonhomogeneous Poisson process.

The failure rate function goes by several aliases: in biostatistics, it is also known as the hazard or hazard rate function; in actuarial science, it is known as the force of mortality or force of decrement; in point process and extreme value theory, it is known as the rate or intensity function; and in economics, its reciprocal is known as Mill's ratio.

The failure rate function is defined as

$$\lambda(t) = \frac{f(t)}{R(t)} \quad t \geq 0.$$

Thus, the failure rate function is the ratio of the probability density function to the reliability function.

All failure rate functions must satisfy two conditions:

$$\int_0^{\infty} \lambda(t)dt = \infty, \ \ \lambda(t) \geq 0 \ \ \text{for all} \ \ t \geq 0. \tag{1}$$

If the failure rate function is known, the cumulative failure rate function, the reliability function, the probability density function, and the mean residual function can be determined by

$$\Lambda(t) = \int_0^t \lambda(u)du$$

$$R(t) = \exp(-\Lambda(t))$$

$$f(t) = \lambda(t)\exp(-\Lambda(t))$$

$$m(t) = \frac{\int_t^{\infty} R(u)du}{R(t)} \quad t \geq 0$$

Also, any one life time distribut0ion representation implies the other four representation (see, for example, Leemis 1995).

A smooth estimation of the failure rate function is a very important topic in both theoretical and applied statistics; Anderson and Senthilselvan(1980) used quadratic spline with discontinuity in the slope at the times of death. O'sullivan (1988) used smoothing splines for the log-hazard function. Senthilselvan(1987) used hyperbolic spline function which is continuous and whose first derivative is discontinuous only at a finite number of points. Kooperberg, et al.(1995) used cubic splines and two additional log terms for log-hazard function. The discussion section of Abrahamowicz, et al.(1992) contains a good review of many of the papers on the use of splines to estimate density and hazard functions in the presence of censored data.

In this paper we estimate failure rate function by a function from a space of cubic splines constrained to be linear (or constant) in tails. Section 2 is devoted to an introduction to spline model for the failure rate function. One advantage of this model is that the representations for all life time distribution representations can be obtained in a closed form. A maximum likelihood estimation procedure is discussed in section 3. Section 4 contains knot deletion procedure. Section 5 contains examples using simulated data.

2. Spline Model for the Failure Rate Function

Let K denote an integer with $K \geq 1$ and let $t_1, \cdots, t_K$ be a (simple) knot sequence in $[0, \infty)$ where $0 < t_1 < \cdots < t_K < \infty$. Let S_0 denote the collection of twice continuously differentiable functions s on $[0, \infty)$ such that the restriction of s to each of the intervals $[0, t_1], [t_1, t_2], \cdots, [t_K, \infty)$ is a cubic polynomial, i.e., s is a polynomial of order 4 (or less) on each of the intervals. Then S_0 is the $(K+4)$-dimensional vector space of cubic splines corresponding to the knot positions $t_1, \cdots, t_K$. Set S denote the subspace of S_0 consisting of the natural cubic splines with knots at $t_1, \cdots, t_K$, i.e., the functions in S that are linear (or constant) on $[0, t_1]$ and $[t_K, \infty)$. This linear vector space is K-dimensional and has a basis $B_1, \cdots, B_K$ of S (see, for example, de Boor 1978 and Schumaker 1981).

Let Θ denote the collection of all column-vector $\theta = (\theta_1, \cdots, \theta_K)^t \in R^K$ such that $\sum_{i=1}^{K} \theta_i B_i(t) > 0$ for all $t > 0$. Given $\theta \in \Theta$, consider the model

$$\lambda(t : \theta) = \sum_{i=1}^{K} \theta_i B_i(t) \qquad t > 0$$

for the failure rate function. For this spline model, the corresponding cumulative failure rate function, reliability function, and probability density function are respectively given by

$$\Lambda(t:\theta) = \sum_{i=1}^{K} \theta_i C_i(t)$$

$$R(t:\theta) = \exp(-\sum_{i=1}^{K} \theta_i C_i(t))$$

$$f(t:\theta) = (\sum_{i=1}^{K} \theta_i B_i(t)) \exp[-\sum_{i=1}^{K} \theta_i C_i(t)]$$

where $C_i(t) = \int_0^t B_i(u)\,du$.

Let X be a random variable having continuous and positive density function f and let $X_1, \cdots, X_n$ be independent random variables having the same distribution as X. The log-likelihood function corresponding to the spline model is defined by

$$l(\theta) = \sum_{i=1}^{n} \log f(x_i : \theta), \quad \theta \in \Theta.$$

The MLE $\widehat{\theta}$ is obtained by maximizing the log-likelihood function. We refer to $\hat{\lambda} = \lambda(\cdot : \widehat{\theta})$ as the spline failure rate estimate.

3. Maximum Likelihood Estimation

Let $I(\theta)$ denote the Hessian of $-l(\theta)$, the $K \times K$ matrix whose (j, k) element is

$$-\frac{\partial^2 l(\theta)}{\partial\theta_j \partial\theta_k} = \sum_{r=1}^{n} \frac{B_j(x_r) B_k(x_r)}{(\lambda(x_r))^2}.$$

It is easily seen that if $\tau \in R^K$, then

$$\tau' I(\theta) \tau = \sum_{j=1}^{n} (\sum_{i=1}^{K} \frac{\tau_i B_i(x_j)}{\lambda(x_j)})^2.$$

Thus $I(\theta)$ is positive definite and hence $l(\theta)$ is strictly concave, the MLE is unique if it exists.

Let $S(\theta)$ be the score function of $l(\theta)$, that is, the K-dimensional vector whose jth element is

$$\frac{\partial}{\partial \theta_j} l(\theta) = \sum_{i=1}^{n} \left(\frac{B_j(x_i)}{\lambda(x_i)} - C_j(x_i) \right).$$

The maximum likelihood equation for $\hat{\theta}$ is $S(\hat{\theta})=0$. We use Newton-Raphson method with step-halving for computing $\hat{\theta}$, to start with an initial guess $\hat{\theta}^{(0)}$ and iteratively determine $\hat{\theta}^{(m+1)}$ by the formula

$$\hat{\theta}^{(m+1)} = \hat{\theta}^{(m)} + \frac{1}{2^M} I^{-1}(\hat{\theta}^{(m)}) S(\hat{\theta}^{(m)})$$

where M is the smallest nonnegative integer such that

$$l\{(\hat{\theta}^{(m)} + \frac{1}{2^M} I^{-1}(\hat{\theta}^{(m)}) S(\hat{\theta}^{(m)})\} \geq l\{(\hat{\theta}^{(m)} + \frac{1}{2^{M+1}} I^{-1}(\hat{\theta}^{(m)}) S(\hat{\theta}^{(m)})\}.$$

We stop the iterations when $l(\hat{\theta}^{(m+1)}) - l(\hat{\theta}^{(m)}) \leq 10^{-6}$.

4. Knot Deletion Procedure

In this section we determine the rules for selecting the number and location of knots. Choosing the number of knots is comparable to choosing a bandwidth in kernel estimation. Too many knots lead to a noisy estimate; too few knots give an estimate that is overly smooth, thereby missing essential details. Knot placement is also important, as many knots may be necessary in an interval where the fluctuation of $\lambda(t)$ is severe. In order to determine the number and location, we can directly apply the stepwise knot deletion method of Smith(1982). First place enough initial knots appropriately and then delete unnecessary knots. According to Stone(1991), for twice continuously differentiable $\lambda(t)$, an optimal rate of convergence $n^{-2/5}$ can be achieved if the number of knots are increased proportionally to $n^{1/5}$. So, we use the integer K closest to $4n^{1/5}$ as a number of initial knots. We will describe an initial knot placement rule: place two knots at the first and the last order statistics and the remaining knots as closely as possible to the equi-spaced percentiles. For example, if the number of initial knots is five, they are placed at the 0, 25, 50, 75, 100 percentiles.

First, consider the problem that the estimates of failure rate function take negative values. The estimates of the failure rate may take negative values for large value of K or on intervals $[0, t_1]$ and $[t_K, \infty)$. So, we used following method to satisfy condition (1).

(i) If $\widehat{\theta}_1(t_1+t_2+t_3)+\widehat{\theta}_2$ less than 0, we let $\theta_1=0$, i.e., the functions are constant on $[0, t_1]$.

(ii) If $\widehat{\theta}_K$ less than 0, we let $\theta_K=0$, i.e., the functions are constant on $[t_K, \infty)$.

(iii) If the minimum value of $\widehat{\lambda}(x_i)$, $i=1, \cdots, n$, is negative, we delete the closest knot to x_{j^*}, where x_{j^*} is argument of minimum value of $\widehat{\lambda}(x_i)$.

Now consider the problem of deleting unnecessary knots among initial knots. Following Smith(1982), the absence of a knot ξ of a spline $\sum_{j=1}^{K}\theta_j B_j(t)$ means that

$$\sum_{j=1}^{K}\theta_j\delta_j(\xi) = 0$$

where $\delta_j(\xi)=B_j^{(3)}(\xi-)-B_j^{(3)}(\xi+)$, $B_j^{(3)}(\xi-)$ and $B_j^{(3)}(\xi+)$ are, respectively, the left-and right-hand limit of $\partial^3 B_j(t)/\partial t^3$ at ξ.

At any step we compute

$$\eta_k = \frac{|\widehat{\psi_k}|}{\mathrm{SE}(\widehat{\psi_k})} \quad k=1, \cdots, K$$

where $\widehat{\psi_k}=\sum_{j=1}^{K}\widehat{\theta_j}\,\delta_j(\xi_k)$ and $\mathrm{SE}(\widehat{\psi_k})=\{\delta^t(\xi_k)(I(\widehat{\theta}))^{-1}\delta(\xi_k)\}^{1/2}$. And we delete the knot having the smallest value of η_k. In this manner, we arrive at a sequence of models indexed by J, which ranges from 0 to K-1. Let $I_L=1$ when the estimated function is constant on $[0, t_1]$, $I_L=0$ otherwise. Let $I_R=1$ when the estimated function is constant on $[t_K, \infty)$, $I_R=0$ otherwise. Let $\widehat{l}_J$ denote the log-likelihood function for the Jth model evaluated at the MLE for that model. Let $\mathrm{AIC}_{\alpha}=-2\,\hat{l}_J+\alpha(K-J-I_L-I_R)$ be the Akaike Information Criterion with parameter penalty α for the Jth model. The traditional value of $\alpha=2$ may lead to a spurious estimate; as a result, Kooperberg and Stone(1991) has suggested $\alpha=3$ and Schwarz(1978) has recommended BIC which is given by AIC_{α} with $\alpha=\log(n)$. According to the limited number of simulations(see section 5), we choose the model corresponding to that value $\hat{J}$ of J that minimizes BIC. This model has $K-\hat{J}$ knots and $K-\hat{J}-I_L-I_R$ free parameters.

5. Examples

The spline failure rate estimation procedure described in Section 2 was applied to Weibull(β,σ), Gamma(β,σ) and Dhillon(β,σ) distributions. The density functions are respectively given by:

$$f(t)=\frac{\beta}{\sigma}\left(\frac{t}{\sigma}\right)^{\beta-1}\exp\left(-\left(\frac{t}{\sigma}\right)^{\beta}\right),$$

$$f(t)=\frac{1}{\Gamma(\beta)}\frac{t^{\beta-1}}{\sigma^{\beta}}\exp\left(-\frac{t}{\sigma}\right),$$

$$f(t)=\sigma\beta(\sigma t)^{\beta-1}\exp((\sigma t)^{\beta})\exp(1-\exp(\sigma t)^{\beta}).$$

The simulation is performed on the subroutine IMSL of the package FORTRAN in SP2 at Seoul National University.

First, to choose the penalty parameter α, we compared Mean Squared Error (MSE)'s of the estimated failure rate functions selected by AIC_{α} with penalty α =2, 3, and $\log(n)$ at each decile t of distribution function F. <Table 1> and 2 respectively show that the results of the simulation from Weibull(β, 1) distribution when β=0.7, 1.0 and 2.0 and Gamma(β, 1) distribution when β=0.8, 1.5 and 2.5 with sample of size n=100, 200 and 500. We replicated 100 times. <Table 1> and 2 illustrate that all estimated failure rate functions selected by BIC have the smallest value of MSE, except in the lower decile of F. So, we select the final spline model by BIC.

< Table 1 > MSE of the estimated function at each decile of Weibull(β,1) distribution

Sample size	$F(t)$	$\beta=0.7$			$\beta=1$			$\beta=2$		
		AIC_2	AIC_3	BIC	AIC_2	AIC_3	BIC	AIC_2	AIC_3	BIC
	0.1	.2111	.1909	.2186	.0304	.0205	.0123	.0476	.0354	.0154
	0.2	.1124	.1168	.1295	.0327	.0146	.0122	.1170	.0912	.0658
	0.3	.0691	.0796	.0916	.0329	.0167	.0141	.0722	.0598	.0408
	0.4	.0431	.0449	.0479	.0290	.0139	.0099	.0729	.0467	.0250
100	0.5	.0310	.0265	.0252	.0305	.0185	.0095	.1551	.0981	.0297
	0.6	.0381	.0326	.0255	.0453	.0322	.0132	.1413	.0985	.0427
	0.7	.0310	.0310	.0296	.0424	.0256	.0150	.2304	.0843	.0473
	0.8	.0217	.0199	.0249	.0509	.0210	.0110	.2605	.1386	.0640
	0.9	.0228	.0217	.0286	.0396	.0192	.0095	.5337	.3651	.2089
	0.1	.1346	.1340	.1349	.0123	.0079	.0055	.0108	.0071	.0029
	0.2	.0893	.0876	.0836	.0104	.0058	.0055	.0137	.0108	.0066
	0.3	.0439	.0465	.0480	.0113	.0060	.0055	.0142	.0093	.0082
	0.4	.0238	.0264	.0278	.0136	.0073	.0055	.0175	.0150	.0098
200	0.5	.0119	.0122	.0105	.0165	.0071	.0055	.0294	.0224	.0131
	0.6	.0167	.0151	.0093	.0140	.0067	.0055	.1218	.1031	.0181
	0.7	.0117	.0111	.0111	.0159	.0060	.0055	.1319	.0697	.0244
	0.8	.0075	.0059	.0076	.0192	.0053	.0055	.1270	.0690	.0328
	0.9	.0140	.0098	.0106	.0174	.0059	.0055	.2266	.1365	.0466
	0.1	.0768	.0765	.0785	.0080	.0065	.0028	.0040	.0027	.0019
	0.2	.0560	.0559	.0549	.0053	.0039	.0028	.0089	.0065	.0046
	0.3	.0286	.0296	.0335	.0054	.0039	.0028	.0086	.0061	.0047
	0.4	.0202	.0202	.0221	.0049	.0036	.0028	.0084	.0057	.0047
500	0.5	.0067	.0064	.0059	.0048	.0033	.0028	.0098	.0072	.0059
	0.6	.0060	.0058	.0058	.0059	.0037	.0028	.0215	.0098	.0075
	0.7	.0046	.0048	.0065	.0070	.0039	.0028	.0248	.0139	.0094
	0.8	.0039	.0031	.0034	.0075	.0041	.0028	.0305	.0196	.0122
	0.9	.0046	.0039	.0040	.0111	.0062	.0028	.0432	.0346	.0173

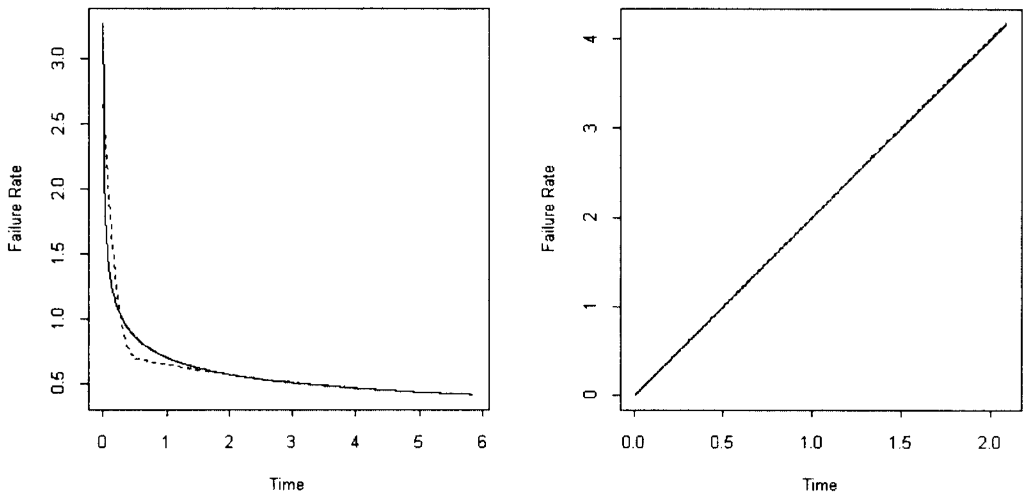
Next, we carried out a simulation study to determine how well the spline failure rate estimate $\hat{\lambda}(t)$ selected by BIC fits the real failure rate function $\lambda(t)$.
In the left side of <Figure 1>, we show the true failure rate function(solid) corresponding to Weibull(0.7,1) distribution. The dotted line corresponds to the estimated failure rate function based on a sample of size 100. The right side of Figure 1 shows the result of a same simulation based on a sample of size 100 from Weibull(2, 1) distribution. From these two examples, we find that spline the failure rate estimator approximates Weibull distributions extremely well for sample of size 100.

< Table 2 > MSE of the estimated function at each decile of Gamma(β,1) distribution

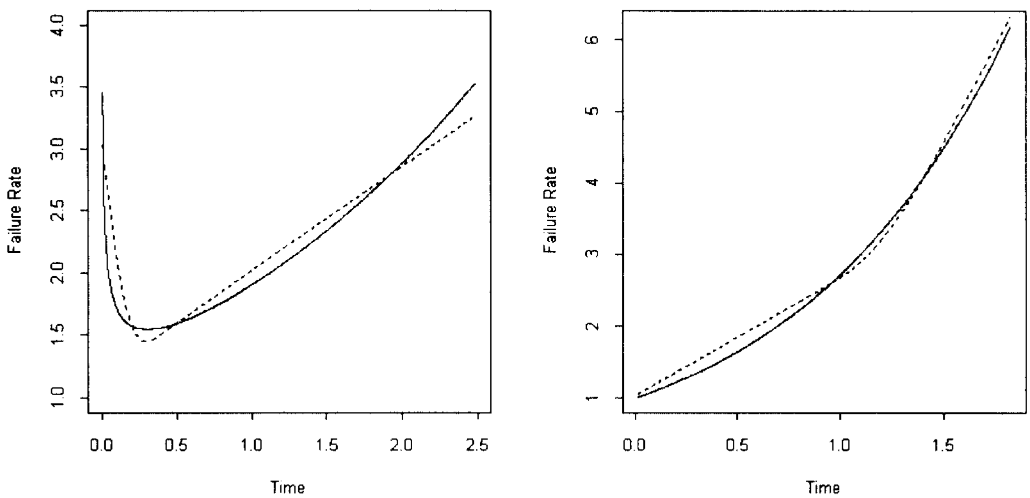
Sample size	$F(t)$	$\beta=0.8$			$\beta=1.5$			$\beta=2.5$		
		AIC_2	AIC_3	BIC	AIC_2	AIC_3	BIC	AIC_2	AIC_3	BIC
100	0.1	.1591	.1617	.1618	.0116	.0143	.0224	.8290	.8286	.8284
	0.2	.0939	.0863	.0718	.0173	.0157	.0108	.0484	.0481	.0471
	0.3	.0902	.0719	.0494	.0165	.0160	.0093	.0145	.0147	.0150
	0.4	.0681	.0410	.0299	.0209	.0169	.0097	.0320	.0327	.0314
	0.5	.0513	.0392	.0290	.0311	.0178	.0126	.0139	.0140	.0136
	0.6	.0828	.0494	.0363	.0307	.0237	.0179	.0118	.0110	.0105
	0.7	.0895	.0512	.0446	.0522	.0356	.0245	.0246	.0206	.0174
	0.8	.0829	.0668	.0511	.0602	.0423	.0351	.0510	.0378	.0379
	0.9	.0890	.0800	.0629	.0800	.0698	.0663	.0640	.0778	.1017
200	0.1	.1173	.1474	.1500	.0062	.0063	.0118	.0027	.0030	.0040
	0.2	.0511	.0582	.0541	.0084	.0085	.0082	.0033	.0041	.0063
	0.3	.0305	.0298	.0241	.0097	.0098	.0090	.0035	.0040	.0057
	0.4	.0232	.0204	.0131	.0113	.0090	.0075	.0038	.0036	.0040
	0.5	.0258	.0158	.0124	.0112	.0099	.0088	.0046	.0042	.0035
	0.6	.0226	.0196	.0185	.0192	.0158	.0137	.0061	.0060	.0044
	0.7	.0312	.0240	.0253	.0185	.0149	.0177	.0098	.0085	.0079
	0.8	.0410	.0283	.0289	.0220	.0183	.0220	.0185	.0151	.0190
	0.9	.0397	.0317	.0341	.0394	.0361	.0349	.0298	.0341	.0569
500	0.1	.0643	.0642	.0852	.0041	.0041	.0049	.0009	.0010	.0012
	0.2	.0375	.0385	.0419	.0050	.0048	.0048	.0013	.0013	.0015
	0.3	.0153	.0153	.0157	.0053	.0055	.0055	.0015	.0014	.0014
	0.4	.0098	.0097	.0081	.0048	.0047	.0038	.0016	.0015	.0014
	0.5	.0097	.0090	.0060	.0055	.0052	.0029	.0018	.0019	.0019
	0.6	.0104	.0092	.0084	.0041	.0039	.0033	.0026	.0025	.0021
	0.7	.0108	.0094	.0106	.0047	.0051	.0041	.0026	.0022	.0024
	0.8	.0127	.0080	.0108	.0063	.0054	.0048	.0034	.0027	.0035
	0.9	.0122	.0112	.0128	.0124	.0092	.0071	.0067	.0060	.0089

<Figure 2> is similar to <Figure 1>, but the underlying distribution for <Figure 2> is Dhillon distribution. The data for the left side of <Figure 2> is from Dhillon(0.7, 1) distribution based on a sample of size 100. In the figure we show the true failure rate function corresponding to this Dhillon distribution together with the estimate for the failure rate function based on the spline model. In the right side of <Figure 2>, we show the result of similar calculation based on a sample of size 100 from Dhillon(1, 1) distribution, i.e., extreme value distribution.
As a result, we have found that the spline failure rate estimate yields a reasonable estimate for the failure rate function.

< Figure 1 > Spline failure rate estimate for Weibull distribution based on sample of size 100. — equals truth and ⋯ equals estimate. Left, β=0.7, σ=1; right, β=2, σ=1.



< Figure 2 > Spline failure rate estimate for Dhillon distribution based on sample of size 100. — equals truth and ⋯ equals estimate. Left, β=0.7, σ=1; right, β=1, σ=1.



Reference

[1] Abrahamowicz, M., Ciampi, A. and Ramsay, J.O.(1992): "Nonparametric Density Estimation for Censored Survival Data : Regression-Spline Approach," *The Canadian Journal of Statistics*, Vol. 20, pp. 171-185.

[2] Anderson, J.A. and Senthilselvan, A.(1980): "Smooth Estimates for the Hazard Function," *Journal of the Royal Statistical Society*, Ser. B, Vol. 42, pp. 322-327.

[3] de Boor, C.(1978): *A Practical Guide to Splines*, Springer-Verlag, New York.

[4] Kooperberg, C., Stone C.J. and Truong, Y.K.(1995): "Hazard Regression," *JASA* Vol. 90, pp. 78-94.

[5] Kooperberg, C. and Stone C.J.(1991): "A Study of Logspline Density Estimation," *Computational Statistics and Data Analysis*, Vol. 12, pp. 327-347.

[6] Leemis, L.M.(1995): *Reliability; Probabilistic Models and Statistical Methods,* Prentice-Hall, New Jersey.

[7] O'sullivan F.(1988): "Fast Computation of Fully Automated Log-Density and Log-Hazard Estimates," *SIAM Journal of Scientific and Statistical Computing*, Vol. 9, pp. 363-379.

[8] Schumaker, L.L.(1981): *Spline functions; Basic Theory*, Wiley, New York.

[9] Schwarz, G. (1978): "Estimating the dimension of model," *Annals of Statistics*, Vol. 6, pp. 461-464.

[10] Senthilelvan, A.(1987): "Penalized Likelihood Estimation of Hazard and Intensity Functions," *Journal of the Royal Statistical Society*, Ser. B, Vol. 49, pp. 170-174.

[11] Smith, P.(1982): "Curve fitting and modeling with splines using statistical variable selection methods," NASA, Langley Research Center, Hampla, VA, NASA Report 166034.

[12] Stone C.J.(1991): Generalized Multivariate Regression Splines, *Technical Report No. 318*, Dept. statist. Univ. California, Berkeley.