Analysis of Censored Data
IMS Lecture Notes – Monograph Series (1995) Volume 27

Maximum Modified Kernel Likelihood Estimation of the Intensity for a Counting Process

P. Anil Kumar and U.V. Naik-Nimbalkar
Department of Statistics
University of Pune, Pune 411007, India

Abstract

A method which may be called a maximum modified kernel likelihood estimation is introduced to estimate the intensity of the multiplicative intensity model for a counting process. This model can be used in survival analysis under general censoring patterns. The asymoptotic properties of the resulting estimator and the selection of the kernel bandwidth are discussed. Its asymptotic distribution is found to be same as that of Ramlau-Hansen's estimator. From a simulation study, it is seen that the finite sample behaviour of the proposed estimator for the hazard rate with right censored data is better than that of Ramlau-Hansen's estimator.

1. Introduction. The multiplicative intensity model introduced by Aalen (1978) is the statistical model for counting processes for which the stochastic intensity admits the decomposition into a functional deterministic factor $\alpha(t)$ and a predictable stochastic process $Y(t)$. This model has been widely applied to the life history data arising in biomedical studies (see Andersen, Borgan, Gill and Keiding (1993)). Using martingales and stochastic integrals, Aalen (1976,1978) developed nonparametric estimators for certain cumulative intensities.

Ramlau - Hansen (1983) proposed an estimator for the intensity of a counting process by smoothing the martingale estimator (the Nelson-Aalen estimator) of the cumulative intensity. Other methods, analogous to that of density estimation, have been studied for estimation of α. Detailed bibliographic remarks concerning these are given in Andersen et al (1993, p.324). It is known that a nonparametric unconstrained maximum likelihood estimator for α does not exist since the likelihood is unbounded. One of the approaches adopted to overcome this problem, is Grenander's (1981) method of sieves. In this approach the log-likelihood function is maximized over a subset of the parameter space with the subset 'converging' to the parameter space as the sample grows. Karr (1987) suggested a sieve using the regularity criteria

$$S(m_n) = \{\alpha \geq 0 : (1/m_n) \leq \alpha \leq m_n, \mid \alpha' \mid < m_n \alpha\},$$

where m_n is a parameter chosen based on the sample and α' is the derivative of α. Later, Leskow and Rozanski (1989) derived a histogram sieve estimator for α. The asymptotic variance of both these estimators is closely related with that of Ramlau-Hansen's estimator.

In this paper we adopt a method which could be considered as a 'local sieve'. We use a kernel function to smooth the likelihood and to obtain an estimate of the intensity. In a recent paper Thavaneswaran and Jagbir Singh (1993) obtained a similar estimator by solving the smoothed optimal estimating function. A similar approach is also mentioned in Hjort (1992).

In Section 2, the construction of a maximum modified kernel likelihood (MMKL) estimator is discussed. Asymptotic properties of the MMKL estimator such as consistency and asymptotic normality are stated in Section 3. Section 4 contains some discussion about the optimum choice of the bandwidth. In the last Section we use this approach to obtain an estimator of the hazard rate function based on right censored data. We compare its finite sample behaviour with that of Ramlau-Hansen's(1983) estimator through simulations.

2. Maximum modified kernel likelihood estimation. Let $(\Omega, \mathcal{F})$ be a measure space equipped with a filtration $\{\mathcal{F}_t, t \in [0,1]\}$ satisfying the *usual conditions.* We observe a sequence of counting processes $\{N^{(n)}(t), t \in [0,1], n \in \mathcal{N}\}$ defined on $(\Omega, \mathcal{F})$ and adapted to the filtration. Let $\mathcal{P}^{(n)} = \{P^n_\alpha; \alpha \in I\}$ be the set of candidate probability measures corresponding to the counting process $N^{(n)}$ where the index set I consists of all left continuous right hand limited, non-negative functions $\alpha \in L_1[0,1]$. We assume that the $(P^n_\alpha, \{\mathcal{F}_t\})$-intensity process of $N^{(n)}$ is given by $\Lambda^{(n)}(t) = \alpha(t) Y^{(n)}(t)$, where $\{Y^{(n)}(t)\}$ is a sequence of $\mathcal{F}_t$- predictable processes which are observable. The Measure P^n is chosen to correspond to the function α equal to one. It is known that for each n, the family is dominated by P^n (Karr (1991) ch. 5) and

$$\frac{dP^n_\alpha}{dP^n} = \exp\left[\int_0^1 (1-\alpha(s))Y^{(n)}(s)ds + \int_0^1 \log(\alpha(s))dN^{(n)}(s)\right].$$

The basis of inference is the log-likelihood function

$$L_n = \int_0^1 (1-\alpha(s))Y^{(n)}(s)ds + \int_0^1 \log(\alpha(s))dN^{(n)}(s).$$

As already observed by Karr (1987) a direct maximization of L_n is meaningless as it is unbounded in α. Heuristically the value of α at a point s

governs the process $N^{(n)}(t)$ only in a neighbourhood of s. Thus in order to estimate $\alpha(s)$ it is enough to consider a portion of the likelihood which gives more emphasis to the behaviour of the process around s. This can be accomplished by computing the kernel log-likelihood

$$L_n(K,h,s) = \int_0^1 \frac{1}{h}K(\frac{s-u}{h})(1-\alpha(u))Y^{(n)}(u)du$$

$$+ \int_0^1 \frac{1}{h}K(\frac{s-u}{h})\log(\alpha(u))dN^{(n)}(u),$$

where K is a non-negative smooth symmetric function having support [-1,1] and satisfying $\int_{-1}^{1} K(u)du = 1$ and h is the bandwidth parameter. However, $L_n(K,h,s)$ is still unbounded if α is allowed to vary in I. If we assume that α is continuous then for small h, $L_n(K,h,s)$ can be approximated by

$$L_n^* = L_n^*(K,h,s) = (1-\alpha(s))\frac{1}{h}\int_0^1 K(\frac{s-u}{h})Y^{(n)}(u)du$$
$$+\log(\alpha(s))\frac{1}{h}\int_0^1 K(\frac{s-u}{h})dN^{(n)}(u). \quad (1)$$

We write $\alpha(s) \equiv a$. Now L_n^* is bounded in 'a' and the maximum of (1) is attained at

$$\hat{a} = \hat{\alpha}_h(s) = \frac{\int_0^1 K((s-u)/h)dN^{(n)}(u)}{\int_0^1 K((s-u)/h)Y^{(n)}(u)du}, \quad (2)$$

which we call the MMKL estimator.

Remark 2.1: The estimator in (2) can also be obtained in a manner analogous to the kernel density estimator discussed in Silverman (1986). More explicitly, a modification of the maximum likelihood histogram sieve estimator

$$\hat{\alpha}^*(s) = \sum_{k=1}^{m} \frac{\int_0^1 I_{A_k} dN}{\int_0^1 I_{A_k} Y^{(n)}(u)du}$$

proposed by Leskow and Rozanski (1989) results in (2). We note that Leskow (1987) has studied a smooth estimator of α given by

$$\hat{\hat{\alpha}}(s) = (1/h)\int_0^1 K((s-u)/h)\hat{\alpha}^*(u)du,$$

whereas Ramlau - Hansen's (1983) estimator of α is

$$\hat{\alpha}^{**}(s) = \frac{1}{h}\int_0^1 K(\frac{s-u}{h})d\hat{\beta}(s),$$

where $\hat{\beta}(s)$ is the martingale estimator (the Nelsen-Aalen estimator) of the cummulative intensity $\int_0^s \alpha(u)$du. Both these are different from the estimator (2).

3. Asymptotic properties. In this section we state the properties of consistency and asymptotic normality of the MMKL estimator. The proofs are omitted as they follow from, by now, standard techniques in this area, namely from the properties of the Martingales arising in the Doob-Meyer decomposition of the counting processes (see Andersen et al (1993)).

We make the following assumptions. Let s be an interior point of $[0,1]$.
A.1 $Y^{(n)}(u)/n \to \xi(u)$ in probability, uniformly in a neighbourhood of s as $n \to \infty$.
A.2 The functions α and ξ are continuous and $\xi(s) > 0$.
A.3 α is differentiable at s.
Consistency:
Theorem 3.1 : Under assumptions A.1 and A.2,
$\hat{\alpha}_h(s) \to \alpha(s)$ in probability as $n \to \infty$, $h \to 0$ and $nh \to \infty$.
Asymptotic Normality:
Theorem 3.2: Assume that A.1 to A.3 hold.
Then, $\sqrt{nh}(\hat{\alpha}_h(s) - \alpha(s))$ converges in distribution to a normal variable with mean zero and variance $(\alpha(s)/\xi(s)) \int_{-1}^{1} K^2(u)du$, as $n \to \infty$, $h \to 0$, such that $nh \to \infty$ and $nh^3 \to 0$.
Remark 3.2 : It may be noticed that the asymptotic distribution of the proposed estimator coincides with that of Ramlau - Hansen's (1983) estimator.

4. Choice of the bandwidth. As in the case of density estimation, the choice of the the bandwidth is crucial. For density estimation cross validation techniques are usually employed. The idea of cross validation is that of minimizing the mean squared error. Below we discuss how an estimate of the mean squared error can be obtained from the sample and thus suggest a method of local bandwidth selection.

We assume that α satisfies the Lipschitz condition of order ρ; that is,

$$| \alpha(s) - \alpha(t) | \leq M | s - t |^{\rho},$$

where M is a constant. In the following expressions, we let

$$K_h = K((s-u)/h).$$

Let

$$\begin{aligned} MSE(\hat{\alpha}_h(s)) &= E(\hat{\alpha}_h(s) - \alpha(s))^2 \\ &= E\left[\frac{\int_0^1 K_h dN^{(n)}(u)}{\int_0^1 K_h Y^{(n)}(u)du} - \alpha(s)\right]^2 \end{aligned}$$

$$\begin{aligned} &= E\left[\frac{\int_0^1 K_h Y^{(n)}(u)(\alpha(u)-\alpha(s))du + \int_0^1 K_h dM^{(n)}}{\int_0^1 K_h Y^{(n)}(u)du}\right]^2 \\ &= I_1 + I_2 + I_3, \end{aligned}$$

where I_1, I_2, and I_3 are defined below and where

$$M^n(t) = N^n(t) - \int_0^1 \alpha(u)Y^n(u)du$$

is a square integrable martingale.

$$I_1 = E\left[\frac{\int_{-1}^1 K(u)Y^{(n)}(s-hu)(\alpha(s-hu)-\alpha(s))du}{\int_{-1}^1 K(u)Y^{(n)}(s-hu)du}\right]^2$$

Using Lipschitz condition,

$$I_1 \le E\left[\frac{\int_{-1}^1 K(u)Y^{(n)}(s-hu)duMh^{\rho}}{\int_{-1}^1 K(u)Y^{(n)}(s-hu)du}\right]^2 = M^2h^{2\rho}.$$

Similarly,

$$\begin{aligned} I_2 &= E\left[\frac{\int_0^1 K_h)dM^{(n)}(u)}{\int_0^1 K_h Y^{(n)}(u)du}\right]^2 \\ &= \frac{1}{nh}E\left[(\frac{1/\sqrt{nh})\int_0^1 K_h dM^{(n)}(u)}{\int_{-1}^1 K(u)(Y^{(n)}(s-uh)/n)du}\right]^2. \end{aligned}$$

Under A.1 and A.2, for large n, I_2 can be approximated by

$$I_2^* = \frac{1}{nh}\frac{\alpha(s)}{\xi(s)}\int_{-1}^1 K^2(u)du.$$

Finally,

$$I_3 = 2E\left[\frac{\int_0^1 K_h Y^{(n)}(u)(\alpha(u)-\alpha(s))du \times \int_0^1 K_h dM^{(n)}(u)}{\left(\int_0^1 K_h Y^{(n)}(u)du\right)^2}\right]$$

Using Cauchy - Schwarz inequality we have

$$I_3 \le 2(I_1I_2)^{1/2}.$$

We propose an estimate $MSE_h^*(s)$ of $MSE(\hat{\alpha}_h(s))$ as follows:

$$\begin{aligned} MSE_h^*(s) &= M^2h^{2\rho} + \frac{1}{nh}\frac{\alpha(s)}{\xi(s)}\int_{-1}^1 K^2(u)du \\ &\quad +2Mh^{\rho}\left(\frac{\alpha(s)}{nh\xi(s)}\int_{-1}^1 K^2(u)du\right)^{1/2} \\ &= \left[Mh^{\rho} + \left(\frac{\alpha(s)}{hn\xi(s)}\int_{-1}^1 K^2(u)du\right)^{1/2}\right]^2. \end{aligned} \quad (3)$$

But in the above expression both $\alpha(s)$ and $\xi(s)$ are unknown. One possible way, is to replace $\alpha(s)$ by its estimator $\hat{\alpha}_h(s)$ and $n\xi(s)$ by $Y_n(s)$. Therefore, in practice one can do the following. Compute

$$CV(h,s) = \left[Mh^\rho + \left(\frac{\hat{\alpha}_h(s)}{hY_n(s)} \int_{-1}^{1} K^2(u)du\right)^{1/2}\right]^2. \tag{4}$$

The optimum choice of h is then the value of h that minimizes $CV(h,s)$. The validity of the above procedure was verified in case of hazard rate estimation through a simulation study.

Remark 4.1: It is easy to see that (3) is minimized at

$$h^* = \left(\frac{1}{2M\rho}\right)^{2/(2\rho+1)} \left(\frac{\alpha(s)}{n\xi(s)} \int_{-1}^{1} K^2(u)du\right)^{1/(2\rho+1)}.$$

Substituting the value of h^* in (3), one can see that

$$\lim_{n\to\infty} \sup_{\alpha\in I(\rho)} \sup_{s\in[0,1]} [MSE_h^*(s)n^{2\rho/(2\rho+1)}] < \infty,$$

where $I(\rho) = \{\alpha \in I : \alpha$ satisfies Lipscitz condition of order $\rho\}$.

Remark 4.2 For $\rho < 1$, the optimal h^* satisfies the conditions $h^* \to 0$, $nh^* \to \infty$ and $nh^{*3} \to 0$ as $n \to \infty$ required for the asymptotic properties.

Remark 4.3: Observe that $h^* = C(1/n\xi(s))^{1/(2\rho+1)}$ and that $Y_n(s)$ can be used as an estimate of $n\xi(s)$. Thus the following subjective choice is appealing. Plot several curves with $h(s) = (1/Y_n(s))^\delta, 0 < \delta < 1$ and choose the estimate that is in close accordance with one's prior idea about the intensity.

5. Hazard rate estimation. Consider independent identically distributed non- negative failure times $X_1, \ldots, X_n$ with distribution function F and hazard rate α. Let $T_1, \ldots, T_n$ be the corresponding censoring times with distribution function H. Assume that the censoring times are independent of the failure times. The number of failures in $[0,t]$, that is,

$$N^{(n)}(t) = \sum_{j=1}^{n} I[X_j \le t, X_j \le T_j]$$

is a counting process with stochastic intensity $\Lambda^{(n)}(t) = \alpha(t)Y^{(n)}(t)$, where $Y^{(n)}(t) = \sum_{j=1}^{n} I[X_j \ge t, T_j \ge t]$ denotes the number of individuals under observation just before time t. The MMKL estimator of $\alpha(s)$ reduces to

$$\hat{\alpha}_h(s) = \frac{\sum_{j=1}^{n} K((s-X_j)/h)D_j}{h\sum_{j=1}^{n}[G(s/h) - G((s-Z_j)/h)]},$$

where D_j is the indicator of death for the j-th individual, $Z_j = \min(X_j, T_j)$ and $G(s)$ satisfies $K(s) = dG(s)/ds$. It may be recalled that the estimator suggested by Ramlau - Hansen (1983) is given by

$$\hat{\alpha}_h^*(s) = \frac{1}{h}\sum_{j=1}^{n} \frac{K((t - X_j)/h)D_j}{Y^{(n)}(X_j)}.$$

Notice that $\hat{\alpha}_h(s)$ makes use of the failure times and the censoring times explictly whereas, Ramlau - Hansen's estimator uses only failure times and the number of uncensored individuals under observation at the time of each failure. Thus maximum modified kernel likelihood estimator makes use of more information available in the data than the Ramlau - Hansen's estimator. In Table 1, we give the results of a simulation study carried out.

Table 1

			CF = .5				CF = .25			
			n=40 h=.3		n=80 h=.2		n=40 h=.3		n=80 h=.2	
C	t	$\alpha(t)$	MSE (1)	MSE (2)	MSE (1)	MSE (2)	MSE (1)	MSE (2)	MSE (1)	MSE (2)
	.3	.913	163	146	141	129	157	152	139	128
	.4	.791	166	153	156	143	161	149	144	136
	.5	.707	172	158	160	152	166	156	153	141
	.6	.645	182	161	166	155	179	159	158	153
.5	.7	.598	201	171	175	164	188	168	166	158
	.8	.559	220	185	189	168	204	182	180	169
	.9	.527	258	220	197	178	221	199	186	178
	.3	.6	43	37	36	28	38	36	32	29
	.4	.8	67	56	45	42	59	51	54	47
	.5	1.0	92	81	61	61	77	68	69	63
	.6	1.2	124	108	94	88	106	101	91	78
2	.7	1.4	189	161	131	122	143	133	130	119
	.8	1.6	282	234	212	193	215	197	186	168
	.9	1.8	519	410	300	281	398	341	344	318
	.3	.27	28	27	21	20	25	23	18	17
	.4	.48	46	42	39	37	41	37	31	29
	.5	.75	63	57	56	51	56	51	44	39
	.6	1.08	113	103	97	92	107	102	76	71
	.7	1.47	176	154	155	146	159	151	113	98
3	.8	1.92	269	238	246	231	252	240	191	165
	.9	2.43	425	391	381	366	408	381	244	211

In the above simulation study, the failure times follow a Weibull distribution with scale parameter 1 and shape parameter C and censoring variable follows a uniform distribution. The simulation is based on 2000 random samples each of size n. Mean squared errors of the estimators are computed at different points for two values of n, censoring fraction CF, bandwidth h and with $C = .5, 2$, and 3. In Table 1, $\alpha(t)$ denotes the hazard rate while MSE(1), MSE(2) denote the mean squared error times 10^3 for the Ramlau-Hansen and the MMKL estimators, respectively. The Epanechnikov kernel is used throughout.

Concluding Remarks:

The method of estimation dicussed above can be used to obtain a smooth estimator for various other models. For example, the method can be used to estimate the drift function of a linear stochastic differential equation when the likelihood can be written using Theorem 7.7 of Liptser and Shiriyayev (1977). The method also can be used to estimate the time dependent covariate effects in a Cox-type regression model discussed by Zucker and Karr (1990) and by Murphy and Sen(1991). Details for the latter model are included in Anilkumar (1994).

References

Aalen, O. (1976) Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3**, 15-27.

Aalen, O. (1978) Nonparametric inference for a family of counting processes.*Ann. Statist.* **6**, 701-726.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993) *Statistical models based on counting processes.* Springer-verlag, New York.

Anilkumar P. (1994) Some contributions to the functional estimation of stochastic processes. Unpublished Ph.D. thesis, Unoversity of Poona.

Grenander, U. (1981) *Abstract Inference*, Wiley, New York.

Hjort N.L. (1992) On Inference in Pararmetric survival data models. *Internat. Statist. Rev.* **60**, 355-387.

Karr, A.F. (1987) Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann. Statist.* **15**, 473-490.

Karr, A.F. (1991) *Point Processes and their Statistical Inference*, Marcel Decker, Inc., New York.

Leskow, J. (1987) Sieve methods in estimation of a periodic intensity of a point process. Unpublished Ph.D. Thesis.

Leskow, J. and Rozanski, R. (1989) Histogram maximum likelihood estimator in the multiplicative intensity model. *Stoch. Proc. Appl.* **31**, 151-159.

Lipster, R.S. and Shiryayev, A.N. (1977) *Statistics of Random Processes Vol. I, General Theory*, Springer Verlag, New York.

Murphy, S.A. and Sen, P.K. (1991) Time dependent coefficients in a Cox-type regression model. *Stoch. Pro. Appl.* **39**, 153-180.

Ramlau-Hansen, H. (1983) Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Thavaneswaran, A. and Jagbir Singh (1993) A note on smoothd estimating functions.*Ann.Inst. Statist. Math.* **45**, 721-729.

Zucker,D. and Karr, A.F. (1990) Nonparametric survival analysis with time-depended covariate effects. A penalized likelihood approach,*Ann. Statist.* **18**, 327-353.