Asymptotic Optimality of the Least-Squares Cross-Validation Bandwidth for Kernel Estimates of Intensity Functions

Maria Mori Brooks and J. Stephen Marron[1]

Department of Statistics, University of North Carolina,

Chapel Hill, NC 27599-3260, USA

November 16, 1989

ABSTRACT

In this paper, kernel function methods are considered for estimating the intensity function of a nonhomogeneous Poisson process. The bandwidth controls the smoothness of the kernel estimator, and therefore, it is desirable to find a data-based bandwidth selection method that performs well in some sense. It will be proven that the least-squares cross-validation bandwidth is asymptotically optimal for kernel intensity estimation.

Keywords: bandwidth selection, intensity function, cross-validation bandwidth, kernel estimation, nonstationary Poisson processes.

[1]Research partially supported by NSF Grant DMS-8902973

1. Introduction

Let $X_1, X_2, \ldots, X_N$ be ordered observations on the interval [0,T] from a nonstationary Poisson process with intensity function $\lambda(x)$. In this paper, we consider estimation of the intensity function $\lambda(x)$. N, the number of observations that occur in the interval [0,T], has a Poisson distribution with $E[N] = \int_0^T \lambda(u)\, du$. See Cox and Isham (1980), Ripley (1981), and Diggle (1983) for further information regarding point processes.

Rosenblatt (1956) introduced a kernel density estimator for independent identically distributed observations, $X_1, X_2, \ldots, X_n$, from an unknown probability distribution f(x). The kernel density estimator is defined as:

$$\hat{f}_h(x) = n^{-1} \sum_{i=1}^{n} K_h(x-X_i) \qquad (1.1)$$

where $K_h(x) = h^{-1} K(x/h)$. The kernel function, K(.), is the "shape" of the weight that is placed on each data point. The smoothing parameter, h, also known as the bandwidth, quantifies the smoothness of $\hat{f}_h(x)$. Using the same motivation, a natural estimate for $\lambda(x)$ is the kernel estimator:

$$\hat{\lambda}_h(x) = \sum_{i=1}^{N} K_h(x-X_i) \qquad x \in [0,T] \qquad (1.2)$$

where $K_h(x) = h^{-1} K(x/h)$. The kernel function is assumed here to be a symmetric probability density function, and h is the smoothing parameter for $\hat{\lambda}_h(x)$. Both of these estimators are appealing since they take on large values in areas where the data are dense and small values where the data are sparse. The kernel density estimator includes a normalization factor, n^{-1}, so that $\hat{f}_h(x)$ is a probability density function; this adjustment is not needed for estimating an intensity function. Theoretical properties of the kernel intensity estimator have been developed by Devroye and Györfi (1984), Leadbetter and Wold (1983), and Ramlau-Hansen (1983a).

The choice of the smoothing parameter is usually far more important than the choice of the kernel function for both kernel density and intensity estimators. See Silverman (1986) for a discussion and examples illustrating this point. Rosenblatt (1971) presented an intuitive description for the tradeoff between smaller and larger values of h. The mean square error (MSE) of $\hat{\lambda}_h(x)$ is the sum of the squared bias and the variance of $\hat{\lambda}_h(x)$. Rosenblatt showed that a small value of h results in high variance; $\hat{\lambda}_h(x)$ is affected by individual observations and hence is more variable. On the other hand, a large value of h results in high bias; $\hat{\lambda}_h(x)$ is very smooth but does not include minor features of the true intensity function. Thus, it is desirable to find a data-based bandwidth that balances the effects of the bias and the variance of the estimate.

Two models are considered here for the intensity function, a simple multiplicative intensity model and a stationary Cox process model. The general multiplicative intensity model, introduced by Aalen (1978), is frequently used to model counting processes. See Anderson and Borgan (1985) for an overview of these models. Diggle (1985) studied the kernel intensity estimator, $\hat{\lambda}_h(x)$, under the stationary Cox process model. He calculated the mean square error of $\hat{\lambda}_h(x)$ using empirical

Bayesian methods and then estimated the "optimal" bandwidth by minimizing an estimate of the MSE over h. When $\hat{\lambda}_h(x)$ is used with a uniform kernel, Diggle and Marron (1988) proved that Diggle's (1985) minimum MSE method and the least-squares cross-validation bandwidth selection method choose the same bandwidth.

For the related setting of kernel density estimation, Hall (1983), Burman (1985), and Stone (1984) have proven that the least-squares cross-validation bandwidth is asymptotically optimal. In this paper, we show that the least-squares cross-validation method is also asymptotically optimal for intensity estimation under both models. In Section 2, we discuss the two intensity models. Section 3 contains the main result regarding the asymptotic optimality of the least-squares cross-validation bandwidth. Finally, the proof of the theorem is presented in Section 4.

2. The Mathematical Models for the Intensity Function

We focus on two mathematical models for the intensity function: a simple multiplicative intensity model and a stationary Cox process model.

The simple multiplicative intensity model is a specific form of Aalen's (1978) multiplicative intensity model. Suppose that $X_1, X_2, \ldots X_N$ are observations from a nonhomogeneous Poisson process with intensity

$$\lambda_c(x) = c\ \alpha(x) \qquad x \epsilon [0,T] \tag{2.1}$$

where c is a constant, and $\alpha(x)$ is an unknown nonnegative deterministic function with $\int_0^T \alpha(x)\,dx = 1$. Given N, the occurrence times, $X_1, X_2, \ldots, X_N$, have the same distribution as the order statistics corresponding to N independent random variables with probability density function $\alpha(x)$ on the interval [0,T]. The kernel estimate of $\lambda_c(x)$ is given in (1.2); under this model, N is a Poisson random variable that has expected value equal to c. It follows from Ramlau-Hansen (1983a) that this estimator is uniformly consistent and asymptotically normal under the simple multiplicative intensity model.

The second model that we consider is the stationary Cox process model. Let $X_1, X_2, \ldots, X_N$ form a realization of a stationary Cox process on [0,T] with intensity function $\lambda_\mu(x)$ where μ is discussed below. A stationary Cox process (also known as a doubly stochastic Poisson process) is defined by :

1) $\{\Lambda(x), x \in \mathbb{R}\}$ is a stationary, non-negative valued random process
2) conditional on the realization $\lambda_\mu(x)$ of $\Lambda(x)$, the point process is a nonhomogeneous Poisson process with rate function $\lambda_\mu(x)$.

See Cox and Isham (1980) for a discussion of doubly stochastic Poisson processes. We also assume that:

3) $E[\Lambda(x)] = \mu$ where μ is a constant ,
4) $E[\Lambda(x)\Lambda(y)] = \upsilon(|x-y|)$ where $\upsilon(x) = \mu^2\,\upsilon_o(x)$ for $\upsilon_o(x)$ a fixed function.

This model is identical to the stationary Cox process model used in Diggle (1985). The kernel estimate of $\lambda_\mu(x)$ in the Cox process model is the same as $\hat{\lambda}_h(x)$ in the multiplicative intensity model for estimating the intensity function from a data set. The difference between the two models is only seen when the estimators are evaluated mathematically. Under the Cox process model, N is a random variable such that $E[N] = E[\int_0^T \Lambda(x)\, dx] = \mu T$.

Asymptotic analysis provides a powerful tool for understanding the behavior of the kernel intensity estimator. Letting $T \longrightarrow \infty$ is not appropriate since this results in all of the new observations occurring at the right endpoint. In the simple multiplicative intensity model, letting $c \longrightarrow \infty$ has the desirable effect of adding observations everywhere on the interval [0,T] and not changing the relative shape of the target function $\lambda_c(x)$ in the limiting process. In other words, $c^{-1}\lambda_c(x)$ is a fixed function as $c \longrightarrow \infty$. Likewise, under the stationary Cox process intensity model, asymptotic results are studied by letting $\mu \longrightarrow \infty$. Consequently, new observations occur over the entire interval [0,T], and assumption 4) above ensures that the relative shape of the curve $\Lambda(x)$ does not change as μ increases.

3. Asymptotic Optimality of the Cross-Validation Bandwidth

For both of the above models, we are interested in finding a data based bandwidth that approximately minimizes the integrated square error (ISE) of $\hat{\lambda}_h$ where

$$\begin{aligned} ISE_\lambda(h) &= \int [\hat{\lambda}_h(x)-\lambda(x)]^2\, dx \\ &= \int_0^T \hat{\lambda}_h(x)^2 dx - 2 \int_0^T \hat{\lambda}_h(x)\, \lambda(x)dx + \int_0^T \lambda(x)^2 dx\ . \qquad (3.1) \end{aligned}$$

For kernel density estimates, Rudemo (1982) and Bowman (1984) suggested using the method of least-squares cross-validation for selecting the bandwidth. In the intensity estimation setting, the cross validation score function is defined as:

$$CV_\lambda(h) = \int_0^T \hat{\lambda}_h(x)^2 dx - 2 \sum_{i=1}^{N} \hat{\lambda}_{hi}(X_i) \tag{3.2}$$

where $\hat{\lambda}_{hi}(x)$ is the leave-one-out estimator,

$$\hat{\lambda}_{hi}(x) = \sum_{\substack{j=1 \\ j\neq i}}^{N} K_h(x-X_j) \ . \tag{3.3}$$

Since $\sum_{i=1}^{N} \hat{\lambda}_{hi}(X_i)$ is a method of moments estimator of $\int_0^T \hat{\lambda}_h(x)\lambda(x)dx$, and $\int_0^T \lambda(x)^2 dx$ is independent of h, $CV_\lambda(h)$ is a reasonable unbiased estimate of the terms in $ISE_\lambda(h)$ that depend on h. Therefore, the bandwidth that minimizes $CV_\lambda(h)$ should be close to the bandwidth that minimizes $ISE_\lambda(h)$.

In the density estimation setting, Hall (1983), Stone (1984) and Burman (1985) proved that when the true density function, f(x), is continuous, the ISE obtained with the cross-validation bandwidth converges almost surely to the minimum ISE. This result gives a sense in which the least-squares cross-validation method is asymptotically optimal for choosing the bandwidth to estimate a density with any amount of underlying smoothness.

Let $\hat{h}_o$ be any bandwidth that minimizes $ISE_\lambda(h)$ and $\hat{h}_{cv}$ any bandwidth that minimizes $CV_\lambda(h)$ (these minima always exist since $ISE_\lambda(h)$ and $CV_\lambda(h)$ are continuous and bounded functions). Assume that :

a) The kernel function, K(.), is a compactly supported probability density function

b) The true intensity function, λ(.), has two continuous bounded derivatives.

c) The bandwidths under consideration come from a set H_c where for each c, $\sup_{h \in H_c} h \leq c^{-\delta}$, $\inf_{h \in H_c} h \leq c^{(-1+\delta)}$, and $\#(H_c) = \{\text{the number of elements in } H_c\} \leq c^{\rho}$ for some constants $\delta>0$ and $\rho>0$. (under the stationary Cox process model, substitute "μ" for "c" in this assumption).

Assumption b) is a common technical assumption which allows Taylor expansion methods to be used for studying the error functions of $\hat{\lambda}_h(x)$. With assumption c), the set of possible bandwidths nearly covers the range of consistent bandwidths. Under these assumptions, the least-squares cross-validation bandwidth is asymptotically optimal for kernel intensity estimation. This result is stated in Theorem 1.

THEOREM 1 : *If assumptions a), b) and c) hold, then, under the simple multiplicative intensity model,*

$$\frac{ISE_{\lambda}(\hat{h}_{cv})}{ISE_{\lambda}(\hat{h}_{o})} \longrightarrow 1 \quad a.s. \quad as\ c \longrightarrow \infty \ ; \tag{3.4}$$

and under the stationary Cox process model,

$$\frac{ISE_{\lambda}(\hat{h}_{cv})}{ISE_{\lambda}(\hat{h}_{o})} \longrightarrow 1 \quad a.s. \quad as\ \mu \longrightarrow \infty \ . \tag{3.5}$$

The mean integrated square error (MISE),

$$MISE_{\lambda}(h) = E[\ \int(\ \hat{\lambda}_h(x) - \lambda(x))^2 dx\ ]\ , \tag{3.6}$$

is another error criterion that is used to evaluate bandwidth selection procedures. Let h_o be the bandwidth that minimizes $MISE_{\lambda}(h)$. By Lemma

1 in Section 4, the ISE and MISE are essentially the same for large c or μ. As a result of Theorem 1, $\hat{h}_{cv}$ is also asymptotically optimal with respect to MISE in the sense that

$$\frac{MISE_\lambda(\hat{h}_{cv})}{MISE_\lambda(h_o)} \longrightarrow 1 \quad \text{a.s.} \quad \text{as } c \longrightarrow \infty \text{ or } \mu \longrightarrow \infty . \tag{3.7}$$

In order to prove Theorem 1, we use arguments similar to the martingale methods employed by Härdle, Marron and Wand (1990) to prove the asymptotic optimality of density derivatives. In addition, a martingale inequality given by Burkholder (1973) is used several times. Details of the proof are presented in Section 4.

4. Proof of Theorem 1

In this section, we outline the proof of Theorem 1. First consider the simple multiplicative intensity model. That is, the underlying intensity function is $\lambda_c(x) = c\,\alpha(x)$, and the kernel intensity estimate is $\hat{\lambda}_h(x) = \sum_{i=1}^{N} K_h(x-X_i)$ where E[N]=c. Assume that assumptions a), b) and c) hold.

Using Taylor expansion methods similar to those in Silverman (1986, p. 39-40), it is straight forward to show that the mean integrated square error (MISE) of $\hat{\lambda}_h$ is

$$MISE_\lambda(h) = h^{-1}c(\int K^2) + h^4c^2\left[\frac{\int u^2K}{2}\right]^2 \int[\alpha''(x)]^2 + o(h^{-1}c+h^4c^2) \tag{4.1}$$

as $h \longrightarrow 0$, $c \longrightarrow \infty$ and $hc \longrightarrow \infty$. Hence, the asymptotic mean integrated square error (AMISE) is

$$\mathrm{AMISE}_\lambda(h) = h^{-1}c\ (\int K^2) + h^4c^2 \left[\frac{\int u^2K}{2}\right]^2 \int[\alpha''(x)]^2 . \qquad (4.2)$$

The two lemmas below are used to prove statement (3.4).

Lemma 1: $\sup_{h\in H_c} \left| \frac{ISE_\lambda(h) - AMISE_\lambda(h)}{AMISE_\lambda(h)} \right| \longrightarrow 0 \quad a.s. \quad as \quad c \longrightarrow \infty .$

Lemma 2: $\sup_{h,b\in H_c} \left| \frac{CV_\lambda(h) - ISE_\lambda(h) - [CV_\lambda(b) - ISE_\lambda(b)]}{AMISE_\lambda(h) - AMISE_\lambda(b)} \right| \longrightarrow 0 \quad a.s.$

$as \ c \longrightarrow \infty .$

Lemma 1 says that the ISE and the AMISE of $\hat{\lambda}_h(x)$ are asymptotically equivalent, and the two lemmas together imply that

$$\sup_{h,b\in H_c} \left| \frac{CV_\lambda(h) - ISE_\lambda(h) - [CV_\lambda(b) - ISE_\lambda(b)]}{ISE_\lambda(h) - ISE_\lambda(b)} \right| \to 0 \quad a.s. \ as \ c \to \infty \qquad (4.3)$$

Since $ISE(\hat{h}_o) \le ISE(\hat{h}_{cv})$ and $CV(\hat{h}_{cv}) \le CV(\hat{h}_o)$, Theorem 1 follows for the simple multiplicative intensity model.

Now, we must prove Lemma 1 and Lemma 2. The details of the proof of Lemma 2 are given below; Lemma 1 is proven using similar martingale methods.

proof of Lemma 2:

Let $g(1,2,\ldots,N) \longrightarrow (1,2,\ldots,N)$ be a random permutation of the numbers $1,2,\ldots,N$. Define $Y_i = X_{g(i)}$. Essentially, the Y_i's are the "unordered" X_i's. Since the X_i's are observations from a nonhomogeneous Poisson process with intensity $\lambda(x)$, the Y_i's are i.i.d. random variables with density $\alpha(x)$. As a result, kernel density methods developed by Härdle, Marron and Wand (1990) can be used to study $\alpha(x)$.

Define:

$$\hat{\alpha}_h(x) = c^{-1}\hat{\lambda}_h(x)$$

$$\hat{\alpha}_{hi}(x) = c^{-1}\hat{\lambda}_{hi}(x)$$

$$CV_\alpha(h) = \int \hat{\alpha}_h(x)^2\,dx - 2c^{-1}\sum_{i=1}^{N}\hat{\alpha}_{hi}(X_i) = c^{-2}CV_\lambda(h) .$$

Hence, $ISE_\alpha(h) = c^{-2}ISE_\lambda(h)$, and $AMISE_\alpha(h) = c^{-2}AMISE_\lambda(h)$. It is not difficult to show that

$$\sup_{h,b\in H_c}\left|\frac{CV_\lambda(h) - ISE_\lambda(h) - [CV_\lambda(b) - ISE_\lambda(b)]}{AMISE_\lambda(h) - AMISE_\lambda(b)}\right|$$

$$= \sup_{h,b\in H_c}\left|\frac{CV_\alpha(h) - ISE_\alpha(h) - [CV_\alpha(b) - ISE_\alpha(b)]}{AMISE_\alpha(h) - AMISE_\alpha(b)}\right|$$

$$\leq 2\sup_{h\in H_c}\left|\frac{c^{-1}\sum_{i=1}^{N}\hat{\alpha}_{hi}(X_i) - \int\hat{\alpha}_h\alpha - c^{-1}\sum_{i=1}^{N}\alpha(X_i) + \int\alpha^2}{AMISE_\alpha(h)}\right|$$

Thus, it suffices to prove the following:

$$\sup_{h\in H_c}\left|\frac{c^{-1}\sum_{i=1}^{N}\hat{\alpha}_{hi}(X_i) - \int\hat{\alpha}_h\alpha - c^{-1}\sum_{i=1}^{N}\alpha(X_i) + \int\alpha^2}{AMISE_\alpha(h)}\right| \to 0 \text{ a.s. as } c\to\infty \quad (4.4)$$

Define:

$$U_{ij} = h^{-1}K((Y_i-Y_j)/h) - h^{-1}\int K((y-Y_j)/h)\,f(y)\,dy - f(Y_i) + \int f^2(y)\,dy$$

$$V_i = E(U_{ij} \mid Y_i)$$

$$W_{ij} = U_{ij} - V_i$$

Note that $E(V_i) = 0$ and $E(W_{ij}|Y_i) = E(W_{ij}|Y_j) = 0$ for $i,j=1,2,\ldots,N$.

For sums over $i=1,2,\ldots,N$, Y_i (the unordered observation) can be replaced by X_i (the ordered observation) in the summand. Moreover, since $N/c = 1 + O_p(c^{-1/2})$, it follows that

$$\lim_{c\to\infty}[c^{-1}\sum_{i=1}^{N} V_i + c^{-2}\sum_{i=1}^{N}\sum_{j\neq i} W_{ij}]$$

$$= \lim_{c\to\infty} [c^{-1}\sum_{i=1}^{N}\hat{\alpha}_{hi}(X_i) - \int\hat{\alpha}_h\alpha - c^{-1}\sum_{i=1}^{N}\alpha(X_i) + \int\alpha^2]$$

In other words, statement (4.4) holds when both (4.5) and (4.6) hold.

$$\sup_{h\in H_c}\left|\frac{c^{-1}\sum_{i=1}^{N} V_i}{\text{AMISE}_\alpha(h)}\right| \longrightarrow 0 \text{ a.s. as } c\longrightarrow\infty \tag{4.5}$$

$$\sup_{h\in H_c}\left|\frac{c^{-2}\sum_{i=1}^{N}\sum_{j\neq i} W_{ij}}{\text{AMISE}_\alpha(h)}\right| \longrightarrow 0 \text{ a.s. as } c\longrightarrow\infty \tag{4.6}$$

Therefore, in order to prove Lemma 2, it is sufficient to prove statements (4.5) and (4.6).

Conditional on N, $\{\sum_{i=1}^{k} V_i\}_{k=1}^{N}$ and $\{\sum_{i=1}^{k}\sum_{j=1}^{i-1} W_{ij}\}_{k=1}^{N}$ are martingales with respect to the σ-fields generated by $\{Y_1,Y_2,\ldots,Y_k\}$. Burkholder's inequality (1973, p.40) implies that for a generic constant A:

$$E[(\sup_{k=1,\ldots N}\sum_{i=1}^{k} V_i)^{2m} \mid N]$$

$$\leq A\, E[(\sum_{i=1}^{N} E[V_i^2])^m \mid N] + A\sum_{i=1}^{\infty} E[|V_i|^{2m} \mid N] \tag{4.7}$$

$$E[(\sup_{k=1,\ldots N}\sum_{i=1}^{k}\sum_{j=1}^{i-1} W_{ij})^{2m} \mid N]$$

$$\leq A\, E[(\sum_{i=1}^{N} E[(\sum_{j=1}^{i-1} W_{ij})^2 \mid Y_1,Y_2,\ldots Y_k])^m \mid N]$$

$$+ A\sum_{i=1}^{\infty} E[|\sum_{j=1}^{i-1} W_{ij}|^{2m} \mid N] \tag{4.8}$$

Since N is a Poisson random variable with expectation equal to c, $E[N^m] \leq A_m c^m$ where A_m is constant for each m. Hence,

$$E[c^{-1}\sum_{i=1}^{N} V_i]^{2m} \le c^{-2m}\, E[\, E[(\sup_{k=1,\ldots N} \sum_{i=1}^{k} V_i\,)^{2m} \mid N\,]]$$

$$\le c^{-2m}\, E[\, A\, N^m\, h^{4m} + A\, N\,] \qquad \text{by (4.7)}$$

$$\le A_m(c^{-m}\, h^{4m} + c^{-2m+1})$$

and

$$E[c^{-2}\sum_{i=1}^{N}\sum_{j=1}^{i-1} W_{ij}]^{2m} \le c^{-4m}\, E[\, A\, N^{2m}\, h^{-m} + A\, N^{m+2}\, h^{-2m}\,] \qquad \text{by (4.8)}$$

$$\le A_m(c^{-2m}\, h^{-m} + c^{-3m+2}\, h^{-2m})$$

For $h\epsilon H_c$, some $\gamma > 0$ and m sufficiently large,

$$E\left[\frac{c^{-1}\sum_{i=1}^{N} V_i}{AMISE_\alpha(h)}\right]^{2m} \le \frac{A_m(c^{-m}h^{4m} + c^{-2m+1})}{(c^{-1}h^{-1} + h^4)^{2m}} \le A_m[h^m + ch^{2m}] \le A_m c^{-\gamma m} \qquad (4.9)$$

Using Chebychev's theorem,

$$\sup_{h\epsilon H_c} P[\ |c^{-1}\sum_{i=1}^{N} V_i| > c^{-\gamma/4}\ AMISE_\alpha(h)\ ] \le A_m\ c^{-(\gamma/2)m} \qquad (4.10)$$

By assumption c), $\#(H_c) \le c^{\rho}$. Choose m such that $m > 2(\rho+2)/\gamma$. Then,

$$\sum_{c=1}^{\infty} P\left[\sup_{h\epsilon H_c}\left|\frac{c^{-1}\sum_{i=1}^{N} V_i}{AMISE_\alpha(h)}\right| > \epsilon\right]$$

$$\le \sum_{c=1}^{\infty} \#(H_c)\ \sup_{h\epsilon H_c} P[\,|c^{-1}\sum_{i=1}^{N} V_i| > \epsilon AMISE_\alpha(h)]$$

$$< \infty$$

Thus, the Borel Cantelli Lemma implies that

$$\sup_{h\epsilon H_c}\left|\frac{c^{-1}\sum_{i=1}^{N} V_i}{AMISE_\alpha(h)}\right| \longrightarrow 0 \text{ a.s. as } c \longrightarrow \infty\ . \qquad (4.11)$$

Following a similar procedure, statement (4.6) can be verified.

Therefore, we have proven Lemma 2.

The proof of Theorem 1 under the Cox process model is parallel to the argument presented above. Define $\alpha(x) = \dfrac{\Lambda(x)}{\int_0^T \Lambda(u)\, du}$; then $\alpha(x)$ is the probability density function of the unordered occurrence times, $Y_1, Y_2, \ldots, Y_N$, when N is known. In addition, define:

$\hat{\alpha}_h(x) = (\mu T)^{-1} \hat{\lambda}_h(x)$, $\hat{\alpha}_{hi}(x) = (\mu T)^{-1} \hat{\lambda}_{hi}(x)$, and

$$CV_\alpha(h) = \int \hat{\alpha}_h(x)^2\, dx - 2(\mu T)^{-1} \sum_{i=1}^{N} \hat{\alpha}_{hi}(X_i) = (\mu T)^{-2} CV_\lambda(h) .$$

Finally, note that N is a Poisson random variable with mean μT, and thus, $E[N^m] \le A_m(\mu T)^m$, $N/(\mu T) = 1 + O_p(\mu^{-1/2})$, and $(\int_0^T \Lambda(x)dx)/(\mu T) = 1 + O_p(\mu^{-1/2})$.

REFERENCES

[1] Aalen, Odd (1978), "Nonparametric Inference for a Family of Counting Processes," The Annals of Statistics, 6, 701-726.

[2] Andersen, Per Kragh and Borgan, Ornulf (1985), Counting Process Models for Life History Data: A Review," Scandinavian Journal of Statistics, 12, 97-158.

[3] Bowman, Adrian W. (1984), "An Alternative of Cross-Validation for the Smoothing of Density Estimates," Biometrika, 71, 363-60.

[4] Burkholder, D. L. (1973), "Distribution Function Inequalities for Martingales," The Annals of Probability, 1, 19-42.

[5] Burman, Prabir (1985), "A Data Dependent Approach to Density Estimation," Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 69, 609-628.

[6] Cox, D. R. and Isham, Valerie (1980), Point Processes, Chapman and Hall, London.

[7] Devroye, Luc and Györfi, Laszlo (1984), Nonparametric Density Estimation: The L_1 View, John Wiley and Sons, New York.

[8] Diggle, Peter J. (1983), Statistical Analysis of Spacial Point Patterns, Academic Press, London.

[9] Diggle, Peter (1985), "A Kernel Method for Smoothing Point Process Data," Applied Statistics, 34, 138-147.

[10] Diggle, Peter and Marron, J. S. (1988), "Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation," Journal of the American Statistical Association, 83, 793-800.

[11] Härdle, Wolfgang, Marron, J. S. and Wand, M. P. (1990), "Bandwidth Choice for Density Derivatives," to appear in Journal of the Royal Statistical Society, Series B.

[12] Hall, Peter (1983), "Large Sample Optimality of Least Squares Cross-Validation in Density Estimation," The Annals of Statistics, 11, 1156-1174.

[13] Leadbetter, M. R. and Wold, Diane (1983), "On Estimation of Point Process Intensities," Contributions to Statistics: Essays in Honour of Norman L. Johnson (P.K. Sen, ed.), North-Holland, Amsterdam, 299-312.

[14] Marron, James Stephen and Härdle, Wolfgang (1986), "Random Approximations to Some Measures of Accuracy in Nonparametric Curve Estimation," Journal of Multivariate Analysis, 20, 91-113.

[15] Ramlau-Hansen, Henrik (1983a), "Smoothing Counting Process Intensities by Means of Kernel Functions," The Annals of Statistics, 11, 453-466.

[16] Ramlau-Hansen, Henrik (1983b), "The Choice of a Kernel Function in the Graduation of Counting Process Intensities," Scandinavian Actuarial Journal, 165-182.

[17] Ripley, B. D. (1981), Spatial Statistics, John Wiley and Sons, New York.

[18] Rosenblatt, Murray (1956), "Remarks on Some Nonparametric Estimates of a Density Function, The Annals of Mathematical Statistics," 27, 832-837.

[19] Rosenblatt, Murray (1971), "Curve Estimates," The Annals of Mathematical Statistics," 42, 1815-1842.

[20] Rudemo, Mats (1982), "Empirical Choice of Histograms and Kernel Density Estimators," Scandinavian Journal of Statistics, 9, 65-78.

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

[22] Stone, Charles (1984), "An Asymptotically Optimal Window Selection Rule For Kernel Density Estimates," The Annals of Statistics, 12, 1285-97.