Penalized maximum likelihood estimation for a function of the intensity of a Poisson point process

Ronaldo Dias, Clécio S. Ferreira and Nancy L. Garcia *

Universidade Estadual de Campinas, Brasil

Abstract

Let $f : [a, b] \to \mathbb{R}$ be an unknown two times differentiable function and consider M to be an α- homogeneous Poisson process on Graf(f). The goal is to estimate f having a sample of the inhomogeneous Poisson process N constructed by dislocating each point of M perpendicularly to Graf(f) by a normal random variable with zero mean and constant variance σ^2. . The exact formulas for the mean measure and the intensity function of N is obtained. Then, the function f is estimated directly using a hybrid spline approach to penalized maximum likelihood. Simulation results indicate the procedure to be consistent as $\alpha \to \infty$ and $\sigma^2 \to 0$.

Key words: nonparametric estimation, H-splines, Poisson processes.

1 Introduction

Consider the following hypothetical situation: a plane dropping leaflets passes through a region A. Suppose that the leaflets land on the ground at a distance to the trajectory followed by the plane which is normally distributed with zero mean and constant

*Address: Departamento de Estatística, IMECC, UNICAMP, Caixa Postal 6065, 13.081-970 - Campinas - SP. `dias@ime.unicamp.br`, `clecio@ime.unicamp.br` and `nancy@ime.unicamp.br`

variance σ^2. Let N be the point process obtained by the position of the leaflets. Studying the process N we want to estimate the trajectory followed by the plane. This problem was studied by Garcia (1995) when the plane follows a linear trajectory. In this case, the intensity of the process N is straightforward and a parametric approach can be used. However, when the trajectory is not linear, it is not so immediate to obtain the intensity of the process. Moreover, the relationship between the intensity of the process and the trajectory is not trivial and a nonparametric approach has to be used.

To be precise, let $f : [a, b] \to \mathbb{R}$ be a two times differentiable function (first derivative f' is absolutely continuous and second derivative f'' is square integrable) and consider $\mathrm{Graf}(f) = \{(v, w); a \le v \le b, w = f(v)\}$ to be the graph of f. We can think of the problem described above as derived from an α- homogeneous Poisson process

$$M = \sum_i \delta_{(V_i, f(V_i))} \tag{1.1}$$

on $\mathrm{Graf}(f)$. Let $\{\varphi_i, i = 1, 2, \ldots\}$ be a sequence of iid zero mean normal random variables with variance σ^2 independent of M. Construct a new process N as

$$N = \sum_i \delta_{(X_i, Y_i)} \tag{1.2}$$

where for $i = 1, 2, \ldots$

$$(X_i, Y_i) := (V_i, f(V_i)) + (-f'(V_i), 1)\frac{\varphi_i}{\sqrt{1 + f'(V_i)^2}}. \tag{1.3}$$

The point process N defined by (1.2) is a inhomogeneous Poisson process in $\mathbb{R}^2$ with mean measure Λ_f and intensity function μ_f which have non-trivial relationship to f (see Theorems 2.1 and 2.2).

In fact, from (1.3) we can see that the relationship $(X, Y) = g(V, \varphi)$ is not one-to-one. Given a function f and a point (X, Y) out of the curve there are several (V, φ) satisfying (1.3). Then, in the non-parametric framework, the latent variables φ are not identifiable. Nevertheless, our interest lies in estimating the function f directly and not through the intensity function.

Estimation of intensity function of Poisson point processes is not a new subject. Kutoyants (1979) considered parametric estimation for univariate intensity function. Ramlau-Hansen (1983) proposed nonparametric estimation through kernel methods. Afterward several authors considered parametric and nonparametric estimation of the intensity function for Poisson processes, for a more detailed account see Kutoyants (1998). However, the problem we focus here has a different flavor, we do not wish to estimate the intensity function μ_f. We wish to estimate the function f and having an estimate of μ_f does not necessarily give a good estimate for f. In this work we propose a modification of the hybrid spline approach to Penalized Likelihood Estimation. Its novelty comes from estimating interactively the two smoothing components and keeping the computational cost low.

Notice that the data are presented as pairs of observations of the form $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$ and it could be erroneously interpreted as a regression problem. This is the wrong approach since the errors are perpendicular to the curve and do not occur only in the Y variable. Therefore, the regression model given by

$$Y_i = f(X_i) + \epsilon_i, \quad i = 1, 2, \ldots, n. \tag{1.4}$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are iid random variables with zero mean is not appropriate. However, it would be interesting to compare the nonparametric regression estimation with the penalized likelihood regression to see if there are some cases where there is not much difference. The computational cost of the regression approach is much less than the maximum likelihood estimation.

Also, the variables φ_i are non-observable and this leads to the fact that given the function f and a point (X, Y) in the process N, there could be several points $(V_i, f(V_i))$ in M such that (1.3) is true. Therefore, this might be viewed as a problem of latent variables but this approach is not going to be pursued in this work.

This paper has two parts: a theoretical and an computational one. First, the theoretical part is developed in Section 2. Theorems 2.1 and 2.2 present the mean measure and intensity function of the process N along with some examples. The proofs of these theorems are given in Appendices 7 and 7. As one can see, the relationship between the function to be estimated and the intensity function is very complex. Section 3 proposes an estimation method based on a spline approach to penalized maximum likelihood estimation. A difficult task is the study of the statistical properties of the proposed estimator. The existent literature deals with estimation of the intensity function and the known results cannot be applied to this proposed estimator. Although no theoretical proofs are given, a simulation study (Section 5) shows that the performance of the algorithm is good, providing much better estimates than a usual non-parametric regression. Moreover, it seems that the estimator is consistent as $\alpha \to \infty$ or $\sigma \to 0$. The analysis of a real dataset consisting of epicenters of earthquakes with magnitude over 5 in Southern California between 1980 and 1998 is given in Section 6.

2 Mean Measure and Intensity

Let $f : [a, b] \to \mathbb{R}$ be a two times differentiable function and consider $\mathrm{Graf}(f) = \{(v, w); a \leq v \leq b, w = f(v)\}$ to be the graph of f. Define an α- homogeneous Poisson process on $\mathrm{Graf}(f)$ as

$$M = \sum_{i=1}^{L} \delta_{(V_i, f(V_i))} \tag{2.1}$$

where L is a Poisson random variable with mean αC, $C = \int_a^b \sqrt{1 + f'(t)^2}dt$ is the length of the curve f in the interval $[a, b]$ and given $L = n$, V_1, V_2,..., V_n are i.i.d.

random variables with common distribution function given by

$$F(x) = \begin{cases} 0 & \text{if } \; x < a \\ \frac{1}{C}\int_a^x \sqrt{1+f'(t)^2}dt & \text{if } \; a \le x < b \\ 1 & \text{if } \; x \ge b \end{cases} \tag{2.2}$$

The proof of this result is immediate and it is left to the reader.

Let $\{\varphi_i, i = 1, 2, \ldots\}$ be a sequence of iid zero mean normal random variables with variance σ^2 independent of M. Construct a new process N as

$$N = \sum_i \delta_{(X_i, Y_i)} \tag{2.3}$$

where for $i = 1, 2, \ldots$

$$(X_i, Y_i) := (V_i, f(V_i)) + (-f'(V_i), 1)\frac{\varphi_i}{\sqrt{1+f'(V_i)^2}}. \tag{2.4}$$

The point process N defined by (2.3) is a inhomogeneous Poisson process in $\mathbb{R}^2$ with mean measure Λ_f and intensity function μ_f depending on f. The following theorems give the relationship between Λ_f, μ_f and f.

Theorem 2.1 *The point process N defined by (2.3) is a inhomogeneous Poisson process in $\mathbb{R}^2$ with mean measure Λ_f determined by*

$$\begin{aligned} \Lambda_f(u,v) \;&:=\; \Lambda_f((-\infty, u] \times (-\infty, v]) \\ &= \frac{\alpha}{\sqrt{2\pi}\sigma}\int_a^b \sqrt{1+f'(x)^2}\left[\int_{-\infty}^{\min(Z_1(u,x), Z_2(v,x))} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x)<0\}} + \right. \\ &\quad \left. + \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x)>0\}} \mathbb{I}_{\{Z_1(u,x)<Z_2(v,x)\}}\right] dx, \;\; u, v \in \mathbb{R} \end{aligned} \tag{2.5}$$

where

$$Z_1(u,x) = \frac{x-u}{f'(x)}\sqrt{1+f'(x)^2} \tag{2.6}$$

and

$$Z_2(v,x) = (v - f(x))\sqrt{1+f'(x)^2}. \tag{2.7}$$

Recall that for a Poisson point process the relationship between the mean measure and its intensity function is given by

$$\Lambda_f(A) := \iint_A \mu_f(x,y)\,dxdy. \tag{2.8}$$

The following definitions and notation will be used to write the intensity function.

Let $a_j \in [a,b]$, $j = 1, ..., J$ be the points in the interval $[a,b]$ such that the derivative of f vanishes, that is, $f'(a_j) = 0$. Define $A_j = (a_j, a_{j+1}]$, $j = 0, ..., J$, with $A_0 = [a, a_1]$ and $A_J = (a_J, b]$, the intervals of $[a,b]$ where f is monotone. Consider

$$h_u(x) := f(x) - \frac{(u-x)}{f'(x)}, \tag{2.9}$$

to be the height of the normal curve to the line at the point $(x, f(x))$ relatively to the axis $x = u$. Figure 2.1 shows the function h_u in the interval $A_j = (0, \pi]$ for values of u to the left, to the right and in the interior of the interval A_j for the particular case $f(x) = \cos(x),\ x \in (0, \pi)$.

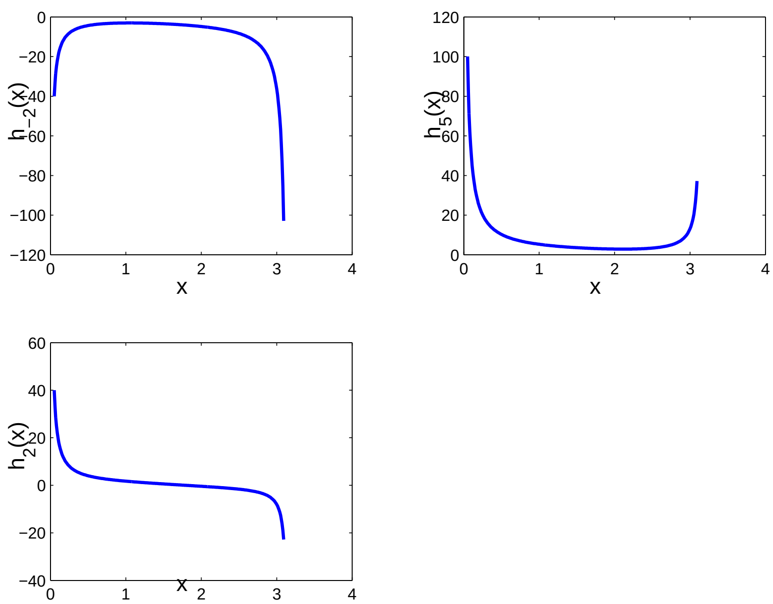

Figure 2.1: Function $h_u(x)$,u=-2,5,2 for $f(x) = \cos(x),\ x \in (0, \pi)$.

Denote $\{x_{j_k}(u,v),\ k=1,...,K_j\}$ the subset of A_j, satisfying

$$h_u(x_{j_k}(u,v)) = v. \tag{2.10}$$

Given the interval A_j, let $A_{j_k}(u,v) = (x_{j_k}(u,v), x_{j_{k+1}}(u,v)]$, $k = 1,...,K_j - 1$, $A_{j_0}(u,v) = (a_j, x_{j_1}(u,v)]$, $A_{j_{K_j}}(u,v) = (x_{j_{K_j}}(u,v), a_{j+1}]$, to be the intervals where h is always bigger or smaller than v. Notice that the set $x_{j_k}(u,v)$ can be the empty set.

Definition 2.1 *Define the following indicator function*

$$\mathbb{I}_{\{h_u(A)<v\}} := \begin{cases} 1, & \text{if } h_u(x) < v, \text{ for all } x \in A \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

The indicator function $\mathbb{I}_{\{h_u(A)>v\}}$ is defined similarly. We also use the notation $\mathbb{I}_{\{x<y\}}$ to indicate that this value is 1 if the inequality is true and 0 otherwise.

Theorem 2.2 *The intensity μ_f of the process N is given by*

$$\begin{aligned}
\mu_f(u,v) \quad = \quad & \frac{\alpha}{\sqrt{2\pi}\sigma}\left\{\sum_{j=1}^{J-1}\left[\sum_{k=1}^{K_j-1}[q_v(x_{j_{k+1}}(u,v)) - q_v(x_{j_k}(u,v))]\times\right.\right. \\
& \left.\left(\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\mathbb{I}_{\{f'(A_j\geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\mathbb{I}_{\{f'(A_j)<0\}}\right)\right]\mathbb{I}_{\{a_j>u\}} + \\
& \left[q_v(x_{j_1}(u,v)) - q_v(x_{j_{K_j}}(u,v)) + \sum_{k=1}^{K_j-1}\left[q_v(x_{j_{k+1}}(u,v)) - q_v(x_{j_k}(u,v))\right] \times\right. \\
& \left.\left(\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\mathbb{I}_{\{f'(A_j)\geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\mathbb{I}_{\{f'(A_j)<0\}}\right)\right]\mathbb{I}_{\{a_{j+1}<u\}} + \\
& \left[q_v(x_{j_1}(u,v)) + \sum_{k=1}^{K_j-1}[q_v(x_{j_{k+1}}(u,v)) - q_v(x_{j_k}(u,v))]\times\right. \\
& \left.\left.\left(\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\mathbb{I}_{\{f'(A_j)\geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\mathbb{I}_{\{f'(A_j)<0\}}\right)\right]\mathbb{I}_{\{a_j<u<a_{j+1}\}}\right\}.
\end{aligned} \tag{2.12}$$

where

$$q_v(x) = \frac{(1+f'(x)^2)e^{-(v-f(x))^2(1+f'(x)^2)/2\sigma^2}}{1+f'(x)^2 - f''(x)(v-f(x))}, \tag{2.13}$$

h_u *is given by (2.9) and* $x_{j_k}(u,v)$ *are solutions of (2.10).*

The proof of Theorem 2.1 and Theorem 2.2 will be presented in Appendices A and B.

2.1 Examples

In this section we present two examples for the mean measure and the intensity function. The first example is the simplest case where f is a straight line passing through the origin. In this case, the expressions can be obtained analytically and it was previously studied by Garcia (1995). The second case we consider $f(x) = \cos(x)$ and the mean measure and intensity function can only be obtained numerically.

Example 2.1 $f(x) = x\tan(\theta)$, *for* $\theta > 0$

First, we notice that (2.10) can be solved explicitly since in this case $x(u,v)$ is the only point in the line f for which its normal line passes through (u,v), represented by P in Figure 2.2.

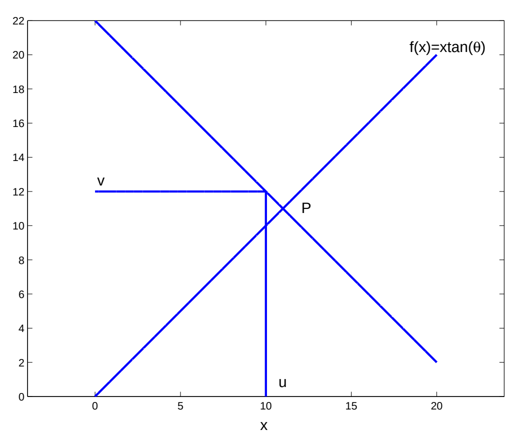

Figure 2.2: x(u,v) for the line $f(x) = x\tan(\theta)$

Therefore,

$$h_u(x(u,v)) = f(x(u,v)) - \frac{(u - x(u,v))}{f'(x(u,v))}$$

which gives the equation

$$x(u,v)\tan\theta - \frac{u - x(u,v)}{\tan\theta} = v$$

with solution

$$x(u,v) = \frac{u + v\tan\theta}{\sec^2\theta}.$$

Second, we find $Z_1(u,x)$ and $Z_2(v,x)$, given by (2.6) and (2.7), respectively.

$$Z_1(u,x) = (x-u)\frac{\sqrt{1+f'(x)^2}}{f'(x)} = (x-u)\frac{\sec\theta}{\tan\theta},$$

$$Z_2(v,x) = (v - f(x))\sqrt{1+f'(x)^2} = (v - x\tan\theta)\sec\theta.$$

It is easy to verify that

$$\frac{\partial}{\partial u}(x(u,v)) = \frac{1}{1+\tan^2\theta} = \frac{1}{\sec^2\theta}.$$

Consequently, by (2.5), we obtain

$$\begin{aligned}
\Lambda_f(u,v) &= \frac{\alpha}{\sqrt{2\pi}\sigma}\int_a^{x(u,v)} \sqrt{1+f'(x)^2}\int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2}dzdx \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma}\int_a^{\frac{u+v\tan\theta}{\sec^2\theta}} \sec\theta \int_{(x-u)\frac{\sec\theta}{\tan\theta}}^{(v-x\tan\theta)\sec\theta} e^{-z^2/2\sigma^2}dzdx.
\end{aligned}$$

The intensity of the process (see (2.12)) is given by

$$\begin{aligned}
\mu_f(u,v) &= \frac{\alpha}{\sqrt{2\pi}\sigma}\frac{(1+f'(x(u,v))^2)e^{-(v-f(x(u,v)))^2(1+f'(x(u,v))^2)/2\sigma^2}}{1+f'(x(u,v))^2 - f''(x(u,v))(v-f(x(u,v)))} \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma}\frac{1}{\sec^2\theta}(1+\tan^2\theta)e^{-\left(v-\frac{u+v\tan\theta}{\sec^2\theta}\tan\theta\right)^2(1+\tan^2\theta)/2\sigma^2} \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma}e^{-(v-(u+v\tan\theta)\sin\theta\cos\theta)^2\sec^2\theta/2\sigma^2} \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma}e^{-(v\cos\theta - u\sin\theta)^2/2\sigma^2}.
\end{aligned} \tag{2.14}$$

Example 2.2 $f(x) = \cos(x)$.

In this case, there is no analytic solutions for the mean measure and intensity function. Figures 2.3 and 2.4 show the numerical solutions. Notice that the intensity function presents bumps inside the local warps.

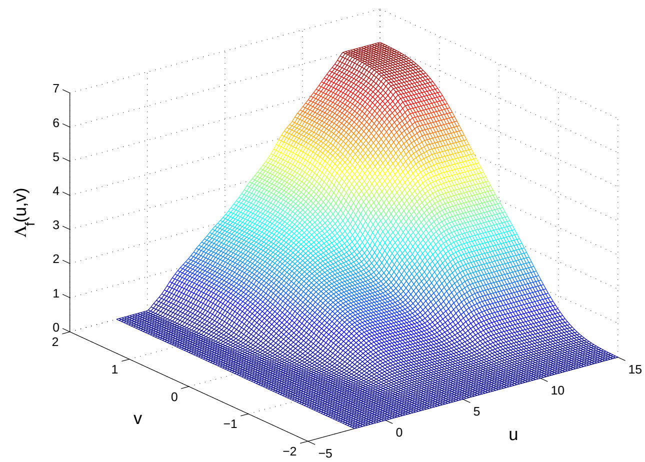

Figure 2.3: Mean measure for N when $f(x) = \cos(x)$

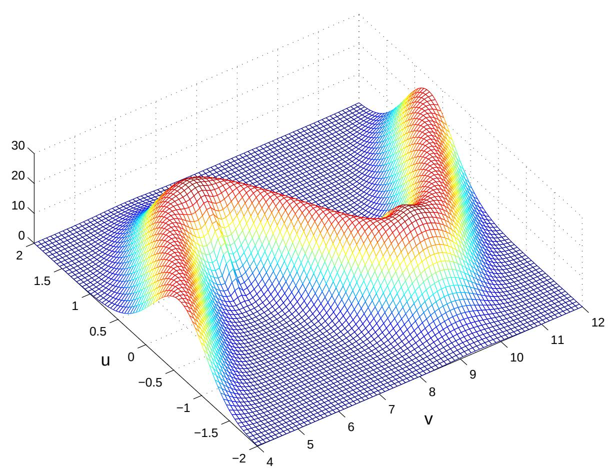

Figure 2.4: Intensity function for N when $f(x) = \cos(x)$

3 Penalized Likelihood Estimation

Let $(X_i, Y_i), i = 1, 2, ..., L$ represent the points of the process N in the set A and μ_f be the intensity function of this process given by (2.12). Let $\mathbf{X} = \{X_i,\ i = 1, ..., L\}$ and $\mathbf{Y} = \{Y_i,\ i = 1, ..., L\}$ be the vector notation of this process. The likelihood function of N is given by (cf. Daley and Vere-Jones (1988))

$$L_A(\mathbf{X}, \mathbf{Y}, f) = \exp\left\{\sum_{i=1}^{L} \log \mu_f(X_i, Y_i) - \int_A (\mu_f(x, y) - 1)dxdy\right\}. \tag{3.1}$$

In this case a measure for goodness of fit can be

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}, f) = -\log(L_A(\mathbf{X}, \mathbf{Y}, f)). \tag{3.2}$$

However, $\mathcal{L}(\mathbf{X}, \mathbf{Y}, f)$ does not have a maximum over the class of all distributions. One possible way to overcome this problem is to consider as a measure for goodness of fit the penalized log-likelihood function. For this let's assume that f is an element of the following functional space $\mathcal{H} = \{h : h' \text{ abs. continuous and } \int (h'')^2 < \infty\}$, and so the penalized log-likelihood function is:

$$\begin{aligned}
\mathcal{L}^*_A(\mathbf{X}, \mathbf{Y}, f; \lambda) &= \log(L_A(\mathbf{X}, \mathbf{Y}, f)) - \lambda \int_a^b (f''(x))^2 dx \qquad (3.3)\\
&= \sum_{i=1}^{L} \log \mu_f(X_i, Y_i) - \int_A (\mu_f(x, y) - 1)dxdy - \lambda \int_a^b (f''(x))^2 dx.
\end{aligned}$$

Therefore, the estimation criterion for this curve is composed by two goals: to maximize the adaptability to the data (likelihood) and to avoid "non-smooth" curves (penalization). To maximize $\mathcal{L}^*_A(\mathbf{X}, \mathbf{Y}, f; \lambda)$ represents a compromise between these two goals.

In order to find the function f that maximizes $\mathcal{L}^*_A(\mathbf{X}, \mathbf{Y}, f; \lambda)$ we are going to use the H-splines method introduced by Dias (1998) in the case of non-parametric density estimation, and by Luo and Wahba (1997) and Dias (1999) in the context of non-parametric regression. The H-splines method combines some features of regression splines and of traditional smoothing splines to obtain a hybrid smoothing procedure which is usually implemented with large data sets and displays a desirable form of spatial adaptability when the underlying function is spatially inhomogeneous in its degree of complexity. Basically, choosing appropriately the number of basis functions, will do most of the work for avoiding over and underfitting. But there is a more important reason why we want to do a penalized approach, namely numerical stability. It is well known that as the number of basis functions increases, the problem becomes more ill-conditioned, which makes the numerical computation less stable. Shortly, we can think of H-splines as having two smoothing components, the number of basis m and the smoothing parameter λ and a trade-off between them. For each λ there is an appropriate m to achieve the best fitting. The problem is how to choose m and λ interactively.

It is well-known that for fixed λ the solution of the optimization problem (3.3), f_λ, is a spline function with knots at the data points (see Wahba (1990)). Let $\mathcal{NCS}$ be the space of the natural cubic splines. Then we can approximate f_λ by $f_{m,\lambda}$ a linear combination of B-splines, that is

$$f_{m,\lambda}(x) = \sum_{i=1}^{m} \beta_i(\lambda) B_{i,4,\xi}(x), \tag{3.4}$$

where $B_{i,4,\xi}(x) =: B_i(x)$ are normalized cubic B-splines with knot sequence ξ and $m \leq n$. The i-th B-spline of order k for the knot sequence ξ is defined by

$$B_i(x) = -(\xi_{k+i} - x_i)[\xi_i, \ldots, \xi_{k+i}](t - \xi_i)_+^{k-1} \quad \text{for all} \quad x \in \mathbb{R},$$

where, $[\xi_i, \ldots, \xi_{k+i}](x - \xi_i)_+^{k-1}$ is $(k-1)$th divided difference of the function $(x - \xi_j)_+^k$ evaluated at points $\xi_i, \ldots, \xi_{k+i}$.

Moreover, B-splines have an important computational property, they are splines which have smallest possible support. In other words, B-splines are zero on a large set. Furthermore, a stable evaluation of B-splines with the aid of a recurrence relation is possible. The B-spline sequence B_j^k consists of nonnegative functions which sum up to 1 and provides a partition of unity.

Therefore, the likelihood $\mathcal{L}_A^*(\mathbf{X}, \mathbf{Y}, f_{m,\lambda}; \lambda)$ given by (3.3) is a now a function of the parameter vector $\beta^T = (\beta_1, \ldots, \beta_m)$ and λ. Thus, for each λ the estimate of the function will be given by

$$\hat{f}_{m,\lambda}(x) = \sum_{i=1}^{m} \hat{\beta}_i(\lambda) B_{i,4,\xi}(x), \tag{3.5}$$

where $\hat{\beta}_i(\lambda)$ (the estimates of β_i for all $i = 1, \ldots, n$) are obtained by solving the numerical problem,

$$\max_{\beta} \sum_{i=1}^{L} \log \mu_{f_{m,\lambda}}(X_i, Y_i) - \int_A (\mu_{f_{m,\lambda}}(x,y) - 1) dx dy - \lambda \int_a^b (f''_{m,\lambda}(x))^2 dx.$$

It remains the problem of finding λ. Usually, the H-splines procedure estimates λ by generalized cross-validation (GCV) method. However, in this particular case, GCV method would introduce a high computational cost to this problem. In order to keep the computational cost at moderate rates we suggest the following modification. At the H-spline step, the estimate of the smoothing parameter λ is obtained in such way that the weight of the smoothness function is the same for the usual Penalized Least Square Criterion and the Penalized Maximum Likelihood Estimation.

4 Computing the estimates

First, let's consider the following modification of the penalized likelihood problem (3.3). Taking $\gamma_1 = \lambda/(1+\lambda)$, we have

$$\begin{aligned} \mathcal{L}_A^*(\mathbf{X}, \mathbf{Y}, f; \gamma_1) &= (1-\gamma_1)\left[\sum_{i=1}^{L} \log \mu_f(X_i, Y_i) - \int_A (\mu_f(x,y) - 1)\, dx\, dy\right] \\ &\quad -\gamma_1 \int_a^b (f''(x))^2 dx \\ &= (1-\gamma_1)\mathcal{L}(\mathbf{X}, \mathbf{Y}, f) - \gamma_1 J(f) \end{aligned} \tag{4.1}$$

where $\mathcal{L}(\mathbf{X}, \mathbf{Y}, f)$ is given by (3.2).

In this case, it is unfeasible to compute λ, or equivalently γ_1, by ordinary cross-validation (CV) or generalized cross-validation (GCV). However, in the regression context GCV not only has very good properties but also it is extremely easy to compute. Taking this fact into account, we suggest to use the penalization given by regression as a starting point for the penalization in the likelihood case. Let's consider the Penalized Least Square Criterion:

$$SP_f(\gamma) = (1-\gamma)SQ + \gamma J(f) \tag{4.2}$$

where $SQ = \sum_{j=1}^{n}(Y_j - f(X_j))^2$, $J(f) = \int_a^b (f''(x))^2 dx$.

Intuitively, we can think of the weight of the penalty function $J(f)$ in the Penalized Least Square (PLS) as

$$\frac{\gamma J(f)}{SP_f(\gamma)} \tag{4.3}$$

while the weight of the penalty $J(f)$ in the Penalized Log-likelihood Equation (PLE) is

$$\frac{\gamma_1 J(f)}{(1-\gamma_1)\mathcal{L}(\mathbf{X}, \mathbf{Y}, f)}. \tag{4.4}$$

Equating (4.3) and (4.4), we get γ_1 as a function of γ as

$$\gamma_1 = \frac{\gamma \mathcal{L}(\mathbf{X}, \mathbf{Y}, f)}{SP_f(\gamma) + \gamma \mathcal{L}(\mathbf{X}, \mathbf{Y}, f)}. \tag{4.5}$$

Notice that $\hat{\gamma}$ (and consequently $\hat{\gamma}_1$) depends on the number of basis functions.

H-splines algorithm for maximizing PLE

1. Initialize the number of basis functions m.

2. Find $\hat{\gamma}_m$ that minimizes

$$GCV(\gamma) = \frac{n^{-1}\sum_{i=1}^{n}(Y_i - f_{m,\gamma}(X_i))^2}{1 - n^{-1}Tr(H(\gamma))},$$

   where $H(\gamma) = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \gamma\Omega)^{-1}\mathbf{X}^T$.

3. Find $\hat{\gamma}_{1m}$ in (4.5) as

$$\hat{\gamma}_{1m} = \frac{\hat{\gamma}_m \mathcal{L}(\mathbf{X}, \mathbf{Y}, f)}{SP_f(\hat{\gamma}_m) + \hat{\gamma}_m \mathcal{L}(\mathbf{X}, \mathbf{Y}, f))}.$$

4. Compute $\hat{f}_{m,\hat{\gamma}_{1m}}$ using (3.5) as

$$\hat{f}_{\hat{\gamma}_{1m}}(x) = \sum_{i=1}^{m} \hat{\beta}_i(\hat{\gamma}_{1m}) B_i(x). \tag{4.6}$$

5. Increase the number of basis functions by one and repeat steps 2 to 4 in order to get $\hat{f}_{m+1,\hat{\gamma}_{1(m+1)}}(x)$

6. Compute $d(\hat{f}_{m,\hat{\gamma}_{1m}}, \hat{f}_{m+1,\hat{\gamma}_{1(m+1)}})$.

7. For a fixed $\delta > 0$, if the distance $d(\hat{f}_{m,\hat{\gamma}_{1m}}, \hat{f}_{m+1,\hat{\gamma}_{1(m+1)}}) < \delta$, stop the procedure. Otherwise go back to 5.

The distance $d(\cdot,\cdot)$ we are using in this work is a pseudo Hellinger distance similar to (Dias, 1999).

In order to simulate the inhomogeneous Poisson process N, we first simulate an α-homogeneous Poisson process on Graf(f) as described in Section 2 and then apply the map (1.3).

The number of initial basis $m \geq 1$ is governed by the prior expectation of the structure of the underlying curve such as maxima, minima, inflection points etc. We suggest the reader to follow some of the rules recommended by Wegman and Wright (1983). These recommendations are based on the assumption of fitting a cubic spline (the most popular case) and are summarized below.

1. Extrema should be center in intervals and inflection points should be located near knot points.

2. No more than one extremum and one inflection point should fall between knots (because a cubic could not fit more).

3. Knot points should be located at data points.

If one follows these recommendations the computation can be reduced substantially. Otherwise, one may start the procedure with one basis function.

5 Simulation Study

For the simulated data, we are going to compute the estimated curves using Penalized Likelihood (PLE) via H-splines as described in Section 4 and, just for comparison, we also show the estimated function using least square regression via H-splines (PLS) (Dias, 1999). Each figure presents the points of the process N, the true curve f, the curves estimated using (PLE) and (PLS). Also, we present, for each approach the affinity derived from the pseudo Hellinger distance between the true function and the estimated one (see, (Dias, 1999) for details), the penalization coefficient λ, the number of knots and the Mean Square Error (MSE) for both estimates given by

$$\mathrm{MSE} = \sum_{i=1}^{L}(y_i - \hat{f}_\lambda(x_i))^2/n. \tag{5.1}$$

Simulations were run in bi-processed Athlon machine with 2.0 GHz processor and 1.5 Gb RAM memory. The software used was Matlab, version 6.1 Release 12, operating in Linux platform.

For the maximization process in PLE, we adopted the following criteria:

- Number of iterations: $300p$, where p is the number of parameters to be estimated;
- Convergence rate: Tolerance 10^{-4} for the parameters and for the likelihood function.
- In order to stop increasing the number of basis functions for the H-spline procedure we required that the affinity between successive estimations would be less than 0.995.

Example 5.1 *The true function $f(x) = x\tan(\theta)$ is a line that passes through the origin.*

In this case, Garcia (1995) computed the closed form for the maximum likelihood estimator of θ which is given by

$$\tilde{\theta} = \frac{1}{2}\arctan\frac{2\sum_A X_iY_i}{\sum_A(X_i^2 - Y_i^2)} + \frac{\pi}{2}\mathrm{sgn}\left(\sum_A X_iY_i\right)\mathbb{I}\left(\sum_A(X_i^2 - Y_i^2) < 0\right)$$

where A is the set where the process is observed. Thus, the parametric estimator is given by $\tilde{f}(x) = x\tan(\tilde{\theta})$.

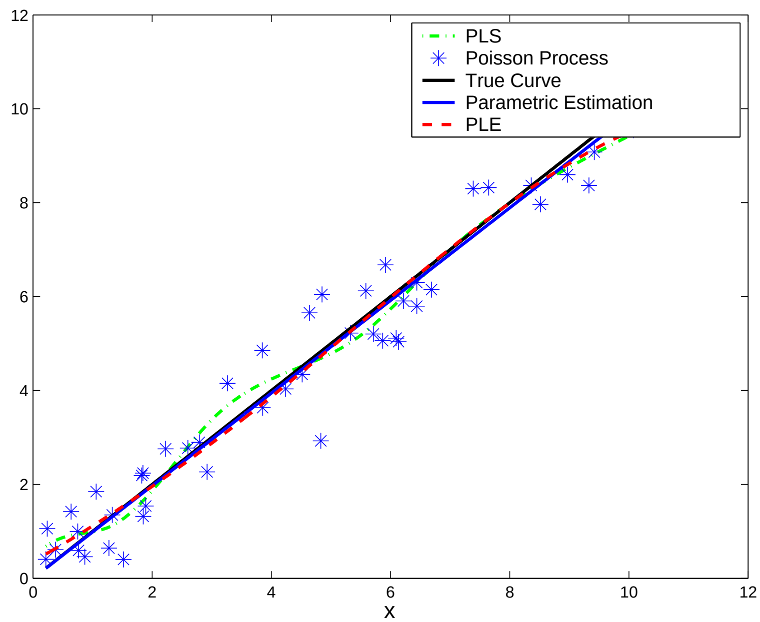

Figure 5.1: Estimated curves when the true function is given by $f(x) = x\tan(\pi/4)$, $x \in (0, 10)$.

In Figure 5.1 and Table 5.1 we can notice that, as expected, the parametric procedure provides a better fitting (MSE = 0.0066). Nevertheless, the PLE estimate is very close to the true curve. On the other hand, the PLS has a poorer performance both in terms of MSE and does not fit a straight line.

Parameters	PLE	PLS
Affinity	-	0.9998
λ	0.001	0.0123
Knots	9	12
MSE	0.0226	0.0666

Table 5.1: Comparative results for the estimation procedures when the true function is a straight line.

Example 5.2 *Here we present several test functions with different degrees of smoothness. The functions used are:* $f_1(x) = \cos 4x$, $f_2(x) = \cos(4\pi x)e^{-x^2/2}$, $f_3(x) = 2 - 5x + 5e^{-100(x-0.5)^2}$ *and* $f_4(x) = 3\ \sin(2\pi x^3)$.

In these cases, the integrals and maximization required in (3.3) cannot be obtained analytically and we are going to use Monte Carlo methods to obtain them.

	f_1		f_2		f_3		f_4	
σ	0.01		0.07		0.04		0.03	
α	12		17		5		8	
$N(A)$	62		86		57		97	
	PLE	PLS	PLE	PLS	PLE	PLS	PLE	PLS
Affinity	0.9954	0.9998	0.9963	0.9966	0.9951	0.9996	0.9963	0.9995
λ	0.0272	0.0039	0.0167	0.0123	0.0012	0.0039	0.0015	0.0015
Number of knots	14	13	20	17	14	15	15	14
MSE	0.0341	0.0583	0.0827	0.1260	0.1464	0.7270	0.1671	0.3980

Table 5.2: Comparative results for the estimation procedures for the functions $f_1(x) = \cos(4x)$, $f_2(x) = \cos(4\pi x)\exp(-x^2/2)$, $f_3(x) = 2 - 5x + 5e^{-100(x-0.5)^2}$ and $f_4(x) = 3\sin(2\pi x^3)$.

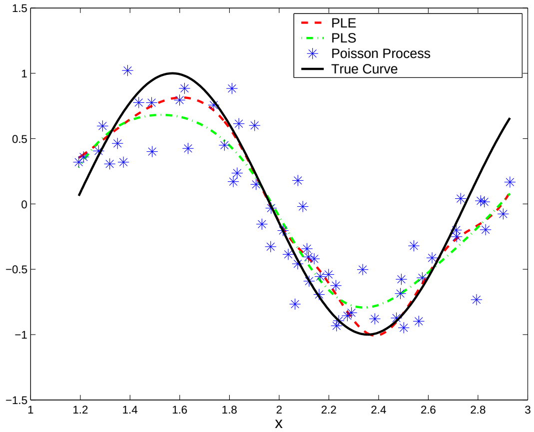

Figure 5.2: Curve estimates for $f_1(x) = \cos(4x),\ x \in (1.2, 2.8)$.

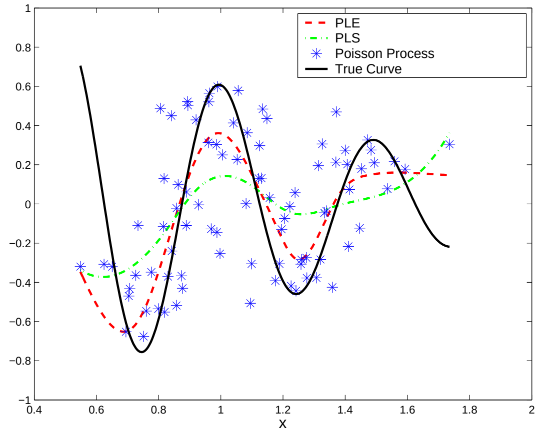

Figure 5.3: Curve estimates for $f_2(x) = \cos(4\pi x)\exp(-x^2/2),\ x \in (0.66, 1.6)$.

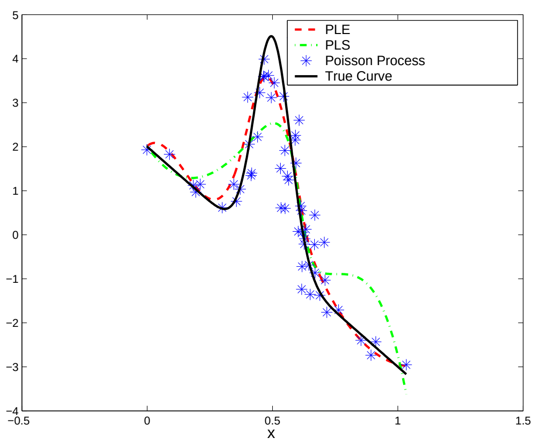

Figure 5.4: Curve estimates for $f_3(x) = 2 - 5x + 5e^{-100(x-0.5)^2},\ x \in (0, 1)$.

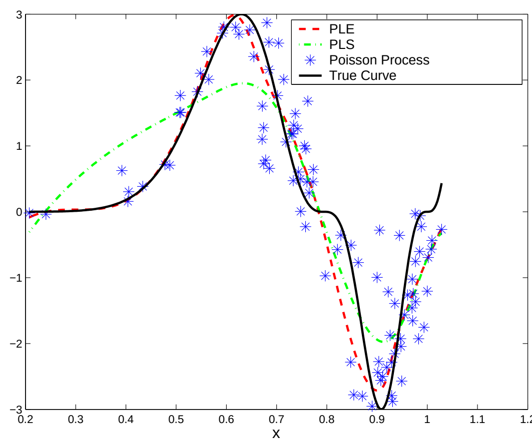

Figure 5.5: Curve estimates for $f_4(x) = 3\sin(2\pi x^3),\ x \in (0.2, 1)$.

By construction of the inhomogeneous Poisson process N, although we have a unidimensional error, it produces error in both coordinates. Looking at Figures 5.2–5.5 we can notice an important characteristic of the process N. It accumulates points inside the curvature of the function f. Therefore, an erroneous model like PLS produces estimates that crosses the true function, that is it goes over the true function when this is concave and under the function when it is convex.

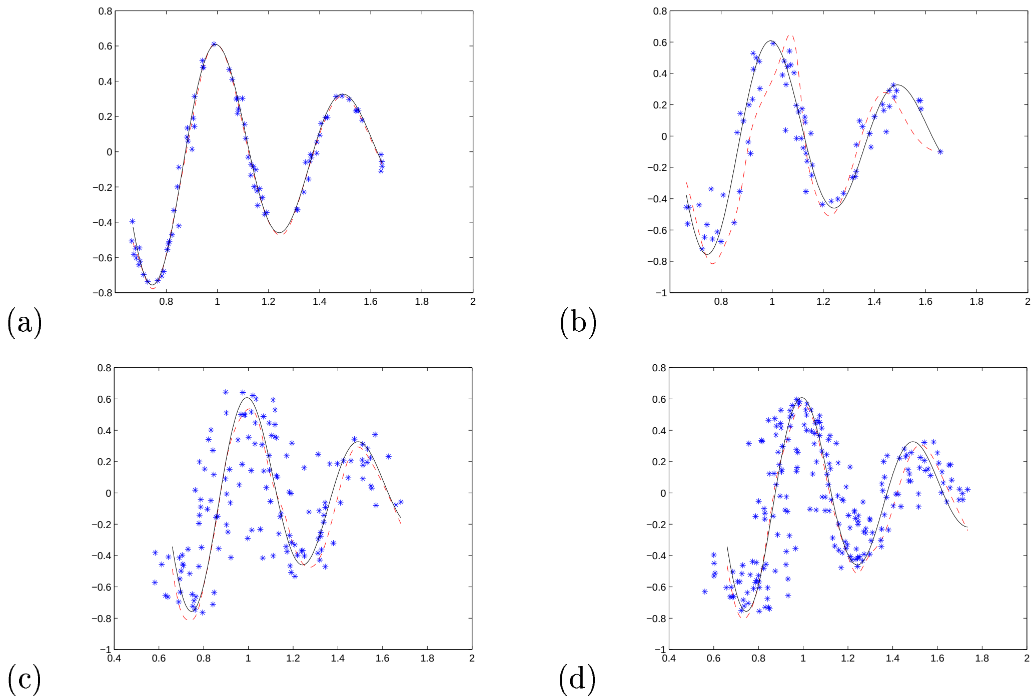

Figure 5.6: PLE curve estimates (dashed) for $f_2(x) = \cos(4\pi x)\exp(-x^2/2)$, $x \in (0.66, 1.6)$ (solid line), (a) $\alpha = 17$, $\sigma = 0.01$, (b) $\alpha = 17$, $\sigma = 0.03$, (c) $\alpha = 35$, $\sigma = 0.07$ (d) $\alpha = 50$, $\sigma = 0.07$.

Consistency. Regarding the consistency of the procedure, Figure 5.6 gives an indication that the estimated curve using PLE is consistent as $\sigma \to 0$ or $\alpha \to \infty$. As it is expected, when σ is very small, the point process falls almost entirely over the curve and the estimation is very good. On the other hand, as α increases we clearly see that PLE approach improves its performance.

The model used for the PLE takes into account the true nature of the errors and as expected it gives much better results. The intensity and mean measure are computed exactly using the function f. In all presented examples we can see that PLE is adequate and fits well. Its main disadvantage is the computational cost. For each iteration of the likelihood, the intensity has to be computed numerically. The function $h(u, x)$ given by (2.9) needs to be computed over the points of the process to be afterward used in (2.10). Just to give a rough idea in order to estimate a curve with m B-spline basis the time elapsed varies between 3600 seg $\times\ m$ and 15000 seg $\times\ m$. This variation depends on the smoothness of the function. Another factor that increases the computational cost is the penalty term in (3.3). In this case, it is unfeasible to compute GCV for PLE since for each point (X_i, Y_i) in the process N we should estimate f taking out this point (leave-one-out method). However, by using the same weight of the smoothness function PLE and PSL the computational cost was substantially reduced.

6 Earthquake data

In this section we apply the PLE to a dataset of 95 earthquakes with magnitude 5 or bigger which occurred in Southern California in the period between 1980 and 1999. This public dataset was obtained from `http://www.scecdc.scec.org/sanandre.html`.

Parameters	PLE	PLS
Affinity	0.9989	0.9999
λ	0.081	0.011
Number of Knots	13	13
MLE	0.2396	0.2280
$\hat{\sigma}$	0.2554	0.2943
$\hat{\alpha}$	12.40	-

Table 6.1: Comparative results for the estimation procedure of earthquake data

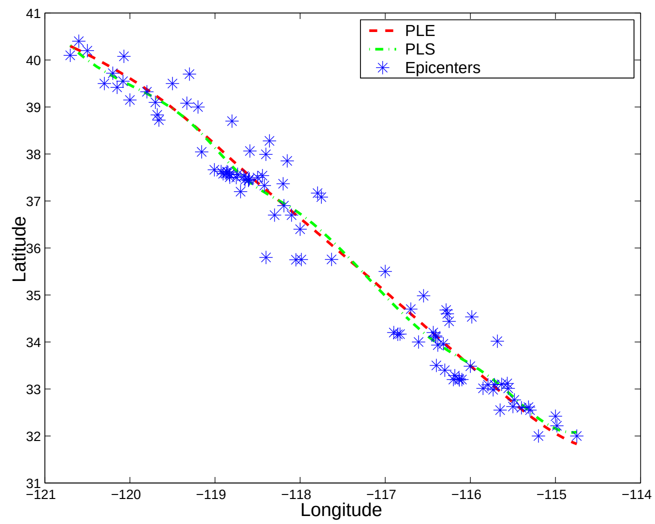

Figure 6.1: PLE and PLS estimated curves for earthquake data

We considered that earthquakes epicenters can be modeled as inhomogeneous Poisson process with distances which are normally distributed and perpendicular to a smooth curve f (maybe a smooth approximation for a geological fault). Since we are taking only high magnitude shocks we can consider that the number of events occurring in disjoint sets are independent random variables.

For this data set, we can use the same procedure as used in the simulation with the additional fact that σ and α are also parameters in the model and have to be estimated. The likelihood function now is $\mathcal{L}_A^*(f;\lambda) = \mathcal{L}_A^*(\beta,\alpha,\sigma;\lambda)$.

As it can be seen from Figure 6.1, our estimate of the geological fault is close to a straight line.

7 Acknowledgments

We would like to thank the anonymous referees for valuable suggestions that greatly improved this paper. This work was partially funded by FAPESP grant 02/01554-5 and CNPq grants 301054/1993-2, 300644/1994-9 and 475763/2003-3.

Appendix A: Proof of Theorem 2.1

In order to find the mean measure we are going to use the fact that a deterministic transformation of a Poisson process is still a Poisson process. Formally,

Proposition 7.1 *Let $T : E_1 \mapsto E_2$ to be a map between Euclidean spaces E_1 and E_2. If N_1 is a Poisson process on E_1 with mean measure Λ_1, then $N_2 = N \circ T^{-1}$ is a Poisson process on E_2 with mean measure $\Lambda_2 := \Lambda_1 \circ T^{-1}$.*

Proposition 7.2 *Suppose that $\{V_n, n = 1, 2, \ldots\}$ are random elements of an Euclidean space E_1 such that*

$$\sum_n \delta_{V_n}$$

is a Poisson process with mean measure Λ. Suppose that $\{\varphi_n, n = 1, 2, \ldots\}$ are random elements of another Euclidean space E_2 with cumulative distribution function F, independent of $\{V_n, n = 1, 2, \ldots\}$. Thus, the point process

$$\sum_n \delta_{(V_n, \varphi_n)}$$

on $E_1 \times E_2$ is a Poisson process with mean measure $\Lambda \times F$, that is for $A_i \subset \mathcal{B}(E_i)$, $i = 1, 2$, we have

$$(\Lambda \times F)(A_1 \times A_2) = (\Lambda \times F)(\{(e_1, e_2) : e_1 \in A_1, e_2 \in A_2\}) = \Lambda(A_1)F(A_2). \tag{7.1}$$

The proof of Propositions 7.2 and 7.1 can be found in Resnick (1992).

Proof of Theorem 2.1

We know that $\{\varphi_i, i = 1, 2, \ldots\}$ are i.i.d. $N(0, \sigma^2)$ random variables and they are independent of the process M. Also, $M' = \sum_n \delta_{V_n}$ is a Poisson process with mean measure

$$\Lambda(A_1) = \alpha \int_{A_1} \sqrt{1 + f'(t)^2}dt$$

and

$$F_\varphi(A_2) = \int_{A_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2} dz.$$

By Proposition 7.2,

$$N_2'(.) = \sum_n \delta_{(V_n,\varphi_n)}(.)$$

is a Poisson process with mean measure

$$\Lambda_1(A_1, A_2) = \alpha \int_{A_1} \sqrt{1+f'(t)^2} dt \int_{A_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2} dz,\ A_1, A_2 \in \mathcal{B}(\mathbb{R}).$$

Consider the following map

$$T(V_n, \varphi_n) = (V_n, f(V_n)) + (-f'(V_n), 1) \frac{\varphi_n}{\sqrt{1+f'(V_n)^2}}. \tag{7.2}$$

Notice that

$$N(A_1 \times A_2) = \sum_n \delta_{T(V_n,\varphi_n)}(A_1 \times A_2). \tag{7.3}$$

Therefore, N is a Poisson process on $\mathbb{R}^2$.

To find its mean measure, consider the family of rectangles

$$\mathbb{A} = \{A(u,v) = (-\infty, u] \times (-\infty, v], u, v \in \mathbb{R}\},$$

which generates the Borel sets on $\mathbb{R}^2$. It is enough to find the mean measure for elements of $\mathbb{A}$.

The inverse image of $A(u,v)$ under T is given by

$$T^{-1}(A(u,v)) = \{x \in [a,b], z \in \mathbb{R} : T(x,z) \in A(u,v)\}.$$

Given $x \in [a,b]$, consider

$$g_x(t) = f(x) - \frac{1}{f'(x)}(t-x), \tag{7.4}$$

to be the line perpendicular to the curve f at the point $(x, f(x))$. Therefore, if V_n is a point of the point process M, then the possible point (X_n, Y_n) corresponding to N will be on the line $g_{V_n}(t)$ (see Figure 7.1).

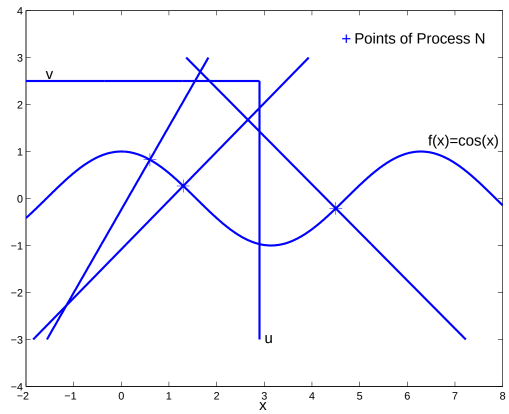

Figure 7.1: The function g_{V_n}, where $M = \sum_n \delta_{(V_n, f(V_n))}$.

According to (1.3),

$$\varphi_n^2 = (Y_n - f(V_n))^2 + (X_n - V_n)^2,$$

that is, φ_n is the Euclidean distance between (X_n, Y_n) and $(V_n, f(V_n))$.

The point where $g_x(t)$ intersects the vertical axis $t = u$ is $(u, f(x) - \frac{u-x}{f'(x)})$ and the point where $g_x(t)$ intersects the horizontal axis $y = v$ is $(x + f'(x)(f(x) - v), v)$. Therefore, $Z_1(u, x)$ and $Z_2(v, x)$, given by (2.6) and (2.7), are the distance between $(x, f(x))$ and $(u, f(x) - \frac{u-x}{f'(x)})$ and $(x, f(x))$ and $(x + f'(x)(f(x) - v), v)$, respectively. Notice that distance here are referred as to be positive or negative according to the orientation to the normal curve the f at the point $(x, f(x))$.

Consequently, we have

$$\begin{aligned} T^{-1}(A(u,v)) &= \{x \in [a,b], z \in \mathbb{R} : T(x,z) \in A(u,v)\} \\ &= \{x \in [a,b], z \in (-\infty, \min\{Z_1(u,x), Z_2(v,x)\}]\mathbb{I}_{\{f'(x)<0\}} \cup \\ &\qquad [Z_1(u,x), Z_2(v,x)]\mathbb{I}_{\{Z_1(u,x)<Z_2(v,x)\}}\mathbb{I}_{\{f'(x)>0\}}\}, \end{aligned}$$

and the mean measure for the process N is given by

$$\begin{aligned} \Lambda_f(u,v) &= \frac{\alpha}{\sqrt{2\pi}\sigma} \int_a^b \sqrt{1 + f'(x)^2} \left[\int_{-\infty}^{\min(Z_1(u,x), Z_2(v,x))} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x)<0\}} + \right. \\ &\left. + \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x)>0\}} \mathbb{I}_{\{Z_1(u,x)<Z_2(v,x)\}} \right] dx. \end{aligned} \tag{7.5}$$

Appendix B: Proof of Theorem 2.2

According to (2.8), the intensity of the process N can be obtained by differentiating the mean measure as

$$\mu_f(u,v) = \frac{\partial^2 \Lambda_f(u,v)}{\partial v \partial u},$$

where Λ_f is given by (7.5).

The following basic theorem allows us to differentiate integrals (see for example Dudewicz and Mishra (1988)):

Theorem 7.1 *(Leibniz's Theorem) If $f(x,\theta)$ is an integrable function with respect to x and $\dfrac{\partial f(x,\theta)}{\partial \theta}$ is a continuous function x and θ, then*

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)\frac{d(b(\theta))}{d\theta} - f(a(\theta),\theta)\frac{d(a(\theta))}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial f(x,\theta)}{\partial \theta}dx. \quad (7.6)$$

Proof of Theorem 2.2 Let $\Lambda_f(u,v)_j$ be the function $\Lambda_f(u,v)$ restricted to the interval A_j.

In order to differentiate $\Lambda_f(u,v)_j$ with respect to the variable v, we need to study not only each case $f'(A_j) > 0$ and $f'(A_j) < 0$ separately, but also, when $a_j > u$, $a_{j+1} < u$ and $a_j < u < a_{j+1}$ since in this case the expressions simplify.

First, notice that $Z_1(u, x_{j_k}(u,v)) = Z_2(v, x_{j_k}(u,v))$, since $h_u(x_{j_k}(u,v)) = v$,

$$f(x_{j_k}(u,v)) - \frac{(u - x_{j_k}(u,v))}{f'(x_{j_k}(u,v))} = v.$$

That is

$$v - f(x_{j_k}(u,v)) = \frac{(x_{j_k}(u,v)) - u}{f'(x_{j_k}(u,v))}. \quad (7.7)$$

Therefore,

$$\begin{aligned} Z_1(u, x_{j_k}(u,v)) &= \frac{x_{j_k}(u,v)-u}{f'(x_{j_k}(u,v))}\sqrt{1+f'(x_{j_k}(u,v))^2} \\ &= (v-f(x_{j_k}(u,v)))\sqrt{1+f'(x_{j_k}(u,v))^2} \\ &= Z_2(v, x_{j_k}(u,v)). \end{aligned}$$

Applying (7.6), we have

Case 1: $a_j > u$

a) If $f'(A_j) > 0$:

$$\frac{\partial}{\partial v}(\Lambda_f(u,v)_j) = \frac{\alpha}{\sqrt{2\pi}\sigma}\sum_{k=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_v(x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v)<v\}}. \quad (7.8)$$

b) If $f'(A_j) < 0$:

$$\begin{aligned} \frac{\partial}{\partial v}(\Lambda_f(u,v)_j) &= \frac{\alpha}{\sqrt{2\pi}\sigma}\left[\frac{\partial}{\partial v}(x_{j_1}(u,v))\sqrt{1+f'(x_{j_1}(u,v))^2}\int_{-\infty}^{Z_1(u,x_{j_1}(u,v))}e^{-z^2/2\sigma^2}dz+\right. \\ &+ \sum_{k=1}^{K_j-1}\left[\frac{\partial}{\partial v}(x_{j_{k+1}}(u,v))\sqrt{1+f'(x_{j_{k+1}}(u,v)^2)}\int_{-\infty}^{Z_1(u,x_{j_{k+1}}(u,v))}e^{-z^2/2\sigma^2}dz+\right. \\ &- \left.\frac{\partial}{\partial v}(x_{j_k}(u,v))\sqrt{1+f'(x_{j_k}(u,v)^2)}\int_{-\infty}^{Z_1(u,x_{j_k}(u,v))}e^{-z^2/2\sigma^2}dz\right]+ \\ &- \frac{\partial}{\partial v}(x_{j_{K_j}}(u,v))\sqrt{1+f'(x_{j_{K_j}}(u,v)^2)}\int_{-\infty}^{Z_1(u,x_{j_{K_j}}(u,v))}e^{-z^2/2\sigma^2}dz+ \\ &+ \left.\sum_{k=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\right]. \\ &= \frac{\alpha}{\sqrt{2\pi}\sigma}\sum_{j=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}. \end{aligned}$$

(7.9)

Case 2: $a_{j+1} < u$

a) If $f'(A_j) \geq 0$:

$$
\begin{aligned}
\frac{\partial}{\partial v}(\Lambda_f(u,v)_j) &= \frac{\alpha}{\sqrt{2\pi}\sigma}\left[\int_{A_{j_0}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx+\right.\\
&+ \sum_{k=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\\
&+ \left.\int_{A_{j_{K_j}}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2)/2\sigma^2}dx+\right].
\end{aligned} \tag{7.10}
$$

b) If $f'(A_j)<0$

$$
\begin{aligned}
&\frac{\partial}{\partial v}(\Lambda_f(u,v)_j) =\\
&\qquad \frac{\alpha}{\sqrt{2\pi}\sigma}\left[\frac{\partial}{\partial v}(x_{j_1}(u,v))\sqrt{1+f'(x_{j_1}(u,v))^2}\int_{-\infty}^{Z_2(v,x_{j_1}(u,v))}e^{-z^2/2\sigma^2}dz+\right.\\
&+ \int_{A_{j_0}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx+\\
&+ \sum_{k=1}^{K_j-1}\left[\frac{\partial}{\partial v}(x_{j_{k+1}}(u,v))\sqrt{1+f'(x_{j_{k+1}}(u,v)^2)}\int_{-\infty}^{Z_2(v,x_{j_{k+1}}(u,v))}e^{-z^2/2\sigma^2}dz+\right.\\
&- \frac{\partial}{\partial v}(x_{j_k}(u,v))\sqrt{1+f'(x_{j_k}(u,v)^2)}\int_{-\infty}^{Z_2(v,x_{j_k}(u,v))}e^{-z^2/2\sigma^2}dz+\\
&+ \left.\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h(A_{j_k}(u,v))>v\}}\right]+\\
&- \frac{\partial}{\partial v}(x_{j_{K_j}}(u,v))\sqrt{1+f'(x_{j_{K_j}}(u,v)^2)}\int_{-\infty}^{Z_2(v,x_{j_{K_j}}(u,v))}e^{-z^2/2\sigma^2}dz+\\
&+ \left.\int_{A_{j_{K_j}}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\right]\\
&= \frac{\alpha}{\sqrt{2\pi}\sigma}\left[\int_{A_{j_0}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx+\right.\\
&+ \sum_{k=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h(A_{j_k}(u,v))>v\}}\\
&+ \left.\int_{A_{j_{K_j}}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\right].
\end{aligned} \tag{7.11}
$$

Case 3: $a_j < u < a_{j+1}$

a) If $f'(A_j) \geq 0$

$$\frac{\partial}{\partial v}(\Lambda_f(u,v)_j) = \frac{\alpha}{\sqrt{2\pi}\sigma}\Bigg[\int_{A_{j_0}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx + \\ + \sum_{k=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\Bigg]. \tag{7.12}$$

b) If $f'(A_j) < 0$

$$\begin{aligned}
&\frac{\partial}{\partial v}(\Lambda_f(u,v)_j) = \\
&\quad \frac{\alpha}{\sqrt{2\pi}\sigma}\Bigg[\frac{\partial}{\partial v}(x_{j_1}(u,v))\sqrt{1+f'(x_{j_1}(u,v))^2}\int_{-\infty}^{Z_2(v,x_{j_1}(u,v))}e^{-z^2/2\sigma^2}dz + \\
&+ \int_{A_{j_0}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx + \\
&+ \sum_{k=1}^{K_j-1}\Bigg[\frac{\partial}{\partial v}(x_{j_{k+1}}(u,v))\sqrt{1+f'(x_{j_{k+1}}(u,v)^2)}\int_{-\infty}^{Z_2(v,x_{j_{k+1}}(u,v))}e^{-z^2/2\sigma^2}dz + \\
&- \frac{\partial}{\partial v}(x_{j_k}(u,v))\sqrt{1+f'(x_{j_k}(u,v)^2)}\int_{-\infty}^{Z_2(v,x_{j_k}(u,v))}e^{-z^2/2\sigma^2}dz + \\
&+ \int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\Bigg] + \\
&- \frac{\partial}{\partial v}(x_{j_{K_j}}(u,v))\sqrt{1+f'(x_{j_{K_j}}(u,v)^2)}\int_{-\infty}^{Z_1(u,x_{j_{K_j}}(u,v))}e^{-z^2/2\sigma^2}dz\Bigg]. \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma}\Bigg[\int_{A_{j_0}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx + \\
&+ \sum_{k=1}^{K_j-1}\int_{A_{j_k}(u,v)}(1+f'(x)^2)e^{-(Z_2(v,x))^2/2\sigma^2}dx\mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\Bigg].
\end{aligned} \tag{7.13}$$

The second derivatives are done in similar way. First, it is necessary to find

$\frac{\partial}{\partial u}(x(u,v))$. From (7.7), we have

$$x(u,v) - u = f'(x(u,v))(v - f(x(u,v))). \tag{7.14}$$

And implicitly differentiating (7.14), it follows

$$\begin{aligned}
\frac{\partial}{\partial u}(x(u,v)) - 1 &= f''(x(u,v))\frac{\partial}{\partial u}(x(u,v))(v - f(x(u,v))) + \\
&- f'(x(u,v))f'(x(u,v))\frac{\partial}{\partial u}(x(u,v)) \\
&= \frac{\partial}{\partial u}(x(u,v))[f''(x(u,v))(v - f(x(u,v))) - f'(x(u,v))^2].
\end{aligned}$$

Obtaining

$$\frac{\partial}{\partial u}(x(u,v)) = \frac{1}{1 + f'(x(u,v))^2 - f''(x(u,v))(v - f(x(u,v)))}. \tag{7.15}$$

Notice that

$$q_v(x(u,v)) = \frac{\partial}{\partial u}(x(u,v))(1 + f'(x(u,v))^2)e^{-(Z_v(x(u,v))^2/2\sigma^2}.$$

Studying each case separately, we finally obtain that the intensity function of the process N is given by

$$\begin{aligned}
\mu_f(u,v) &= \frac{\alpha}{\sqrt{2\pi}\sigma}\Bigg\{\sum_{j=1}^{J-1}\Bigg[\sum_{k=1}^{K_j-1}[q_v(x_{j_{k+1}}(u,v)) - q_v(x_{j_k}(u,v))]\times \\
&\Big[\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\mathbb{I}_{\{f'(A_j\geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\mathbb{I}_{\{f'(A_j)<0\}}\Big]\Bigg]\mathbb{I}_{\{a_j>u\}} + \\
&\Bigg[q_v(x_{j_1}(u,v)) - q_v(x_{j_{K_j}}(u,v)) + \sum_{k=1}^{K_j-1}\left[q_v(x_{j_{k+1}}(u,v)) - q_v(x_{j_k}(u,v))\right]\times \\
&\Big[\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\mathbb{I}_{\{f'(A_j)\geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\mathbb{I}_{\{f'(A_j)<0\}}\Big]\Bigg]\mathbb{I}_{\{a_{j+1}<u\}} + \\
&\Bigg[q_v(x_{j_1}(u,v)) + \sum_{k=1}^{K_j-1}[q_v(x_{j_{k+1}}(u,v)) - q_v(x_{j_k}(u,v))]\times \\
&\Big[\mathbb{I}_{\{h_u(A_{j_k}(u,v))<v\}}\mathbb{I}_{\{f'(A_j)\geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u,v))>v\}}\mathbb{I}_{\{f'(A_j)<0\}}\Big]\Bigg]\mathbb{I}_{\{a_j<u<a_{j+1}\}}\Bigg\}.
\end{aligned}$$

References

Daley, D. J. and Vere-Jones, D. (1988). *An introduction to the theory of point processes*, Springer-Verlag, New York.

Dias, R. (1998). Density estimation via hybrid splines, *Journal of Statistical Computation and Simulation* **60**: 277–294.

Dias, R. (1999). Sequential adaptive non parametric regression via H-splines, *Communications in Statistics: Computations and Simulations* **28**: 501–515.

Dudewicz, E. J. and Mishra, S. N. (1988). *Modern mathematical statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Garcia, N. L. (1995). Maximum likelihood estimation and suboptimal stopping time for a Poisson point process, *Rebrape* **9**(1): 67–82.

Kutoyants, Y. A. (1979). Intensity parameter estimation of an inhomogeneous Poisson process, *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.* **8**(2): 137–149.

Kutoyants, Y. A. (1998). *Statistical inference for spatial Poisson processes*, Vol. 134 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Luo, Z. and Wahba, G. (1997). Hybrid adaptive splines, *Journal of the American Statistical Association* **92**: 107–116.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions, *Ann. Statist.* **11**(2): 453–466.

Resnick, S. (1992). *Adventures in stochastic processes*, Birkhäuser Boston Inc., Boston, MA.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM:PA.

Wegman, E., J. and Wright, I. W. (1983). Splines in statistics, *Journal of the American Statistical Association* **78**: 351–365.