Semidefinite and Second Order Cone Programming Seminar
Fall 2012
Project Report
Shape Constraints And Its Application

Instructor: Farid Alizadeh
Student Name: Joonhee Lee

12/3/2012

1 Overview

In this project, I will introduce a method to estimate the function by cubic splines. As an example, I will focus on the non-homogeneous Poisson process. Since this method uses an optimization model based on the maximum likelihood principle, it may applied to other stochastic processes. The most critical point of this presentation is a nonnegativity constriant on the splines. In addition, choosing the proper number of spline knots m can be done by the Cross-Validation technique.

2 Goal and Motivation

Let $\mathbb{H}$ be a inifinite dimensional linear space. And the goal is to estimate unkonwn real-valued function $f \in \mathbb{H}$ based on a finite number of observations $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N$ and its response $y(\mathbf{t}_1), y(\mathbf{t}_2), \ldots, y(\mathbf{t}_N)$. That is,

$$f(t) = \sum_{i=0}^{\infty} a_i e_i(t)$$

where e_i's are bases of the infinite dimensional space.
This function can be a loss function, density function, or the arrival rate. The key idea of this project is that the shape constraint is involved in the prior assumption of the true function.

Our goal is to estimate the infinite number of a_i's. However, since it is impossible in practice to estimate infinite dimensional object, the reduced dimensional problem must be considered. Let

$$\mathbb{H}_0 \subseteq \mathbb{H}_1 \subseteq \mathbb{H}_2 \subseteq \cdots \mathbb{H}$$

and each $\mathbb{H}_i$ is a fininite dimensional linear space. That is, $\dim(\mathbb{H}_i) = n_i$.
For example, let's consider for $f \in \mathbb{H}_1$ case.

$$f(t_1) = \sum_{i=0}^{n_1} a_i e_i(t)$$

Here, e_i's are bases of the n_1-dimensional linear space.
It is important to take an account that an estimated function from $\mathbb{H}_1$ maybe too simple. When the (oversimplified) estimated function may not accurately generalize the true function f, it is called under-fitting. On the other hand, if the dimesion is too high, then a function (over) fits data along with the "noise" of data. Therefore, the balance between under-fitting and over-fitting must be considered.

The goal is to present an introductory yet thorough investigation on the shape constrainted estimation using nonnegative cubic splines. In some researches, it has been generalized to any degree spline with nonnegativity constraint is enough.

3 Preliminaries

Before I discuss about the shape constraint, some of preliminary knowledge are useful to know.

Definition 1 (Non-homogeneous Poisson Process)

$$\mathbb{P}[(N(b) - N(a)) = k] = \frac{e^{-\lambda_{a,b}}(\lambda_{a,b})^k}{k!}, \quad \textit{where } \lambda_{a,b} = \int_a^b \lambda(t)dt$$

Therefore, the arrival rate may change over time. Note that the homogeneous Poisson process is just a special case of the non-homogeneous one. Since the rate changes over time, the most reasonable guess for estimating rate function, $\lambda(t)$, is a spline.

Definition 2 (Spline) *A spline is a sufficiently smooth polynomial function that is piecewise-defined, and possesses a high degree of smoothness at the places where the polynomial pieces connect (which are known as knots).*

That is,

$$\lambda(t) = p^{(i)}(t) = \sum_{t=0}^{k} p_l^{(}i)(t - a_{i-1})^t \quad \forall\, t \in [a_{i-1}, a_i]$$

For the simplicity, I will focus on the *nonnegative cubic spline* (k=3) and equidistant knots.

Definition 3 (MLE) *For each sample point $\boldsymbol{x}$, let $\hat{\theta}$ be a parameter value at which $L(\theta\,|x)$ attains its maximum as a function of θ, with x held fixed.*
A maximum likelihood estimator (MLE) of the parameter θ based on a sample $\boldsymbol{X}$ is $\hat{\theta}(\boldsymbol{X})$.

The maximum likelihood estimator suggests that the parameter value can be estimated from the observation. Similarly, if the maximum likelihood prinicible is adopted, the parameter *function* can be estimated. To ensure MLE be nonnegative, we need to employ a notion of linear functional operator.

Definition 4 (Linear functional operator) *For $f(\cdot) \geq 0$ throughout $[\underline{x}, \bar{x}]$, some linear functional operator L has a general form $Lf(\cdot) \geq 0$.*

- L to be the identity mapping
- Let f(·) be non-decreasing, let L = D to be the differential operator
- Let f(·) be convex, let $L = D^2$ to be th second derivative operator.

Definition 5 (Polynomials)

Univariate polynomial :

$$p(t) = p_0 + p_1 t + p_2 t^2 + \cdots + p_{2d} t^{2d} \qquad \forall t \in \mathbb{R}$$

Multivariate polynomial (Cone of nonnegative polynomial):

$$P_{n,d}[t_1, \ldots, t_d] = \left\{ P(t_1, \ldots, t_n) \quad | \quad \begin{array}{l} P \textit{ is a polynomial} \\ P(t_1, \ldots, t_n) \geq 0 \qquad \forall t \in \mathbb{R} \end{array} \right\}$$

Let $P_{n,d}$ be a positive semidefinite polynomials. And this is a convex cone. Define sum of squares (SOS) polynomial :

$$\sum_{n,d}[t_1, \ldots, t_d] = \left\{ P(t_1, \ldots, t_n) \quad | \quad \begin{array}{l} P \textit{ is a polynomial and } \exists\, P_1, P_2, \ldots, P_n \\ \textit{such that } P = P_1^2 + \cdots + P_n^2 \end{array} \right\}$$

- For univariate polynomial P, P is also sum of squares.
- $\sum_{n,d} \subset P_{n,d}$

In general, $f \in P_{n,d}$ does not imply $f \in \sum_{n,d}$. However, $f \in \sum_{n,d}$ implies $f \in P_{n,d}$. It is critical that testing if $f \in P_{n,d}$ is NP-hard while testing if $f \in \sum_{n,d}$ is an "easy" problem.

4 Constraints

4.1 Periodicity

Periodicity is an important feature. Consider a model without periodicity. Then, for example, in estimating arrival rate, it is impossible to conclude that the current estimation will work for the future interval. Furthermore, in some interval, no arraive may observed in some intervals. Then, in that interval, any function may fit and it leads to the poor estimation of true function. However, when the periodicity is assumed, every intervals have a sufficient data as the study considers more and more periods.

4.2 Smoothness

Simple piecewise-constant and piecewise-linear models of spline provides a fast computing time along with arbitrarily close approximation when m is sufficiently large. However, those methods tend to produce estimates of $\lambda(\cdot)$ having abrupt changes in the arrival rate or its derivative. There are some situations in which non-smooth forms of $\lambda(\cdot)$ is appropriate; but, in other cases, smooth form will provide more realistic models.

• In this project, $\lambda(\cdot)$ assumed to have a periodicity and smoothness. That means, throughout time interval, $t \in [0, cT]$, must be smooth. Recall the definition of splines :

$$\lambda(t) = p^{(i)}(t) = \sum_{t=0}^{k} p_l^(i)(t - a_{i-1})^t \quad \forall\, t \in [a_{i-1}, a_i]$$

By periodicity, $\lambda(t) = \omega(t(\mathrm{mod}T))$ for $0 = t_0 < t_1 < t_2 < \ldots < t_m = T < t_{m+1} < \ldots < t_{cm} = cT$. In order to obtain $\lambda(\cdot)$ continuous and twice differentiable (for cubic splines), it is necessary to have $\omega(0) = \omega(T), \omega'(0) = \omega'(T), \omega''(0) = \omega''(T)$. That is, for each period, the first knot and the last knot must be continuous and twice differentiable in order to get a smooth $\lambda(\cdot)$.

4.3 Shape Constraints

To estimate the unknown function $f(\cdot)$, imposing shape constraints trims the searching space. For example, the radius of a tree is monotonically increasing as its age increases. Then, it is natural to assume nonnegavity in the rate of growth of tree.
Another example is an arrival rate. At a restaurant, customers are coming while it is open. In the worst case, there is no customer coming during certain hours. However, it is impossible to have a negative number of customers. Then, non-nagative rate of arrival rate of customers is a natural assumption.
In this point of view, shape constrainted approach is somewhere in between non-parametric and parametric.

5 Model Formulation

5.1 Non-periodic Case

In this presentation, the objective functions are based on the maximum likelihood principle. More specifically, the maximum likelihood functions come from the nonhomogeneous Poisson model. That is,

$$f(\mathbf{t},\lambda) = \prod_{j=1}^{n} f_j(t_j,\lambda) = \prod_{j=1}^{n} \lambda(t_j)\exp\Big(-\int_{t_{j-1}}^{t_j}\lambda(t)dt\Big)$$

And, we assume $-\int_{t_n}^{\bar{t}}\lambda(t)dt = 0$. To utilize this information,

$$\tilde{f}(\mathbf{t},\lambda) = \exp\Big(-\int_{t_n}^{\bar{t}}\lambda(t)dt\Big)f(\mathbf{t},\lambda)$$

$$L(\mathbf{t},\lambda) = \ln(\tilde{f}(\mathbf{t},\lambda)) = \sum_{j=1}^{n}\ln\lambda(t_j) - \int_{\underline{t}}^{\bar{t}}\lambda(t)dt$$

5.2 Nonnegativity Constraint

Recall that the objective function is maximizing the likelihood function I discussed in the previous section. For some subintervals, without nonnegativity, $\int_{t_i}^{t_{i+1}}\lambda(t)dt$ tends to become a subzero function. More specifically, for some subintervals observed with zero arrival, the function tries to go below zero. Let the feasible region is a cone $\bar{\Lambda}$ for $\lambda(\cdot)$. Nonnegativity is a crucial constraint of $p_l^{(i)}$ in order to get a nonnegative arrival rate function throughout $[\underline{t}, \bar{t}]$. Without the nonnegativity constraint, the likelihood function corresponding to a cubic spline arrival rate may be unbounded. To see this, assume :

$$\exists\bar{\mu}(\cdot)\in\bar{\Lambda}\ :\ \bar{\mu}(t_j) > 0 \quad \forall j = 1,2,\dots,n, \quad \int_{\underline{t}}^{\bar{t}}\bar{\mu}dt \le 0.$$

Since $\bar{\Lambda}$ is a cone, $\lambda(\cdot) = \alpha\bar{\mu}(\cdot)$ also lies in a cone. Then, $L(\mathbf{t},\lambda) = \sum_{j=1}^{n}\ln\lambda(t_j) - \int_{\underline{t}}^{\bar{t}}\lambda(t)dt$ becomes arbitrarily large as α increases. Thus, in order to obtain a well=defined problem, nonnegativity constraint is required.

5.3 Periodic Case

For periodic case, $\lambda(t) = \omega(t(\mathrm{mod}T))$ for $0 < t_1 < t_2 < \ldots < t_n < \ldots < cT$ for a positive integer c. And period T > 0, and $\omega : [0,T) \to \mathbb{R}$. Rewrite the loglikelihood function in terms of $\omega(\cdot)$,

$$\lambda(t,\omega) = \sum_{j=1}^{n}\ln\omega(t_j(\mathrm{mod}T)) - c\int_0^T\omega(t)dt.$$

- Combining a periodic cubic spline and loglikelihood function above,

$$L(t,p) = \sum_{j=1}^{n} \ln\Big[\sum_{t=0}^{3} p_l^{(i_j)}(t_j(\mathrm{mod}T) - a_{i_j-1})^t\Big] - c\sum_{i=1}^{m}\sum_{t=0}^{3} p_l^{(i)}\frac{d_i^{l+1}}{l+1}$$

- Now, the only missing information is the size of m.

6 Determining m

As it is discussed in the beginning, too small m does not fully exploit the available data, whereas too large m results in overfitting. Among many model selection procedure, k-fold cross-validation is studied.
1. choose small number of m, say m_1
2. choose a random subset of arrival times T and set them aside
3. Solve the problem
4. Evaluate the likelihood function value for the data from step 2.
This procedure will be repeated R times. Finally, average value of $\bar{L}(m_1)$ of the likelihood function over this sample. Then, for the next step, choose $m_2 > m_1$ and so on.
At the end, the maximum among $\bar{L}(m_1), \bar{L}(m_2), \ldots, \bar{L}(m_M)$ is selected with the optimal size of m.

References

[1] Alizadeh, F., Goldfarb, D., *Second-Order Cone Programming*, Mathematical Programming manuscript, 2010.

[2] Alizadeh, F., Eckstein, J., Noyan, N., and Rudolf, G. Arrival rate approximation by nonnegative cubic splines. *Operations Research*, 56:140-156, 2008

[3] Papp, D., Alizadeh, F., *Shape constrained estimation using nonnegative splines*, 2012

[4] Casella, G. and Berger, R. L., *Statistical Inference*, Duxbury Press, 2010.

[5] Ross, Sheldon M., *Simulation*, Academic Press, 2006