A Modeling and Data Analysis of Laser Beam Propagation in the Maritime Domain

Midshipman 1/C Benjamin C. Etringer

Adviser: Professor Reza Malek-Madani, Mathematics Department

Laser Background

Random processes occur throughout nature. The first step to understanding the statistics behind the event is to compute an accurate probability density function for the data collected from the event. However, for stochastic processes, there is no way to compute the exact probability density function. Therefore, we will use different methods to compute probability density functions from given stochastic data. The stochastic data that we will utilize are laser data that have already been collected. We will use the laser data to evaluate each method for computing the approximate probability density function.

Project Background

Laser data exhibits stochastic behavior when propagated through the maritime domain. We would like to compute an approximate probability density function to help us better understand the impact the atmosphere has on the laser in the maritime domain. We will use three methods: (1) Naïve/Kernel Method, (2) Barakat Method through lower-order moments, and (3) Gaussian Mixture Techniques.

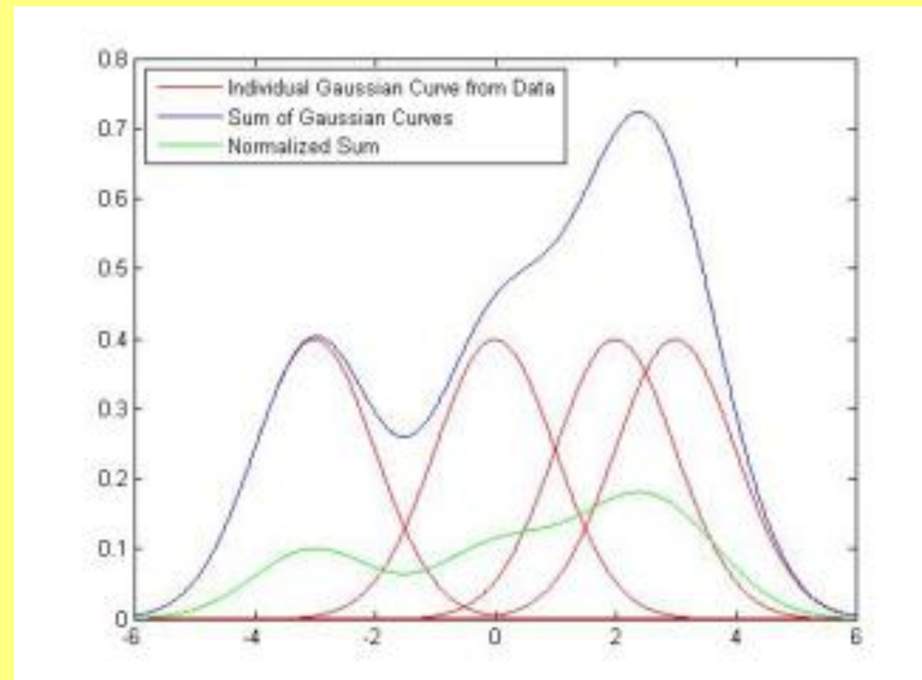
Kernel Method

The Kernel Method is described by Silverman. It is a mixture technique in which a known probability density function, that is dependent upon a single data point, is computed for each data point. The collection of probability density functions is then summed together and normalized to create a probability density function with appropriate area equal to 1. For the Kernel Method with the Gaussian curve as a mixture:

$$p_k(x) = \frac{1}{N}\sum_{i=1}^{N} G_i$$

$$G_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{e\sigma^2}}$$

$$\mu = x_i \qquad \sigma = \frac{\Delta_x}{2\sqrt{N}}$$



Gaussian Mixture Method

The Gaussian Mixture Method (GMM) is a clustering technique that is remarkably similar to the Kernel Method. With the Kernel Method, there are N Gaussian Curves for N data points. GMM lets the user input the number of clusters to increase computational efficiency. The method assumes that each cluster is uniformly spaced on the domain of the data points. The points are readjusted iteratively until they converge.

$$p_{GMM}(x) = \frac{1}{N}\sum_{i=1}^{N} W_i\, G_i \qquad G_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{e\sigma^2}} \qquad \sigma = \frac{x_{i+1} - x_{i-1}}{2}$$

$$\mu = x_i \qquad W_i = \frac{n}{N}$$

Barakat Method

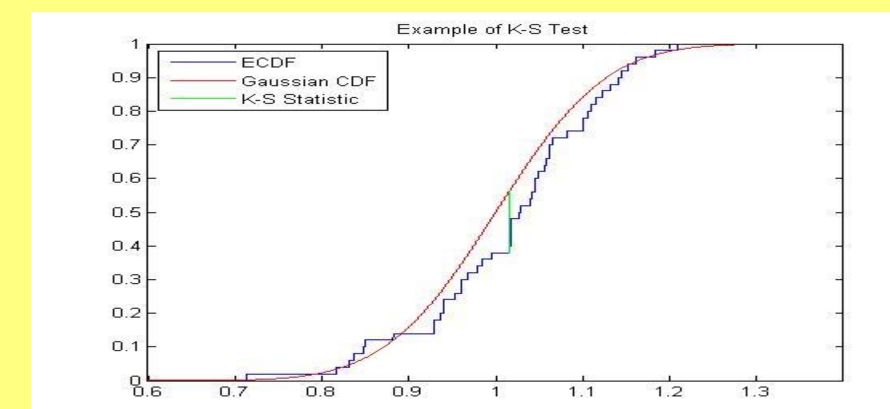
Barakat argues that the first five moments of *h* are sufficient to approximate the PDF of *h* reasonably well. The method involves the use of Generalized Laguerre Polynomials, which are described as L_N here. We intend to evaluate this assumption for the laser data that exhibits significant noise once propagated in the maritime domain.

$$L_N^{\beta-1}(x) = \sum_{n=0}^{N} \binom{N+\beta-1}{N-n} \frac{(-x)^n}{n!} \qquad W(h) = W_g(h)\sum_{n=0}^{\infty} W_n L_n^{\beta-1}\frac{\beta h}{\mu}$$

$$W_N = N!\,\Gamma(\beta)\sum_{n=0}^{N} \frac{\left(-\frac{\beta}{\mu}\right)^n < h^n >}{n!\,(N-n)!\,\Gamma(\beta+n)} \qquad W_g(h) = \frac{1}{\Gamma(\beta)}\left(\frac{\beta}{h}\right)^{\beta} h^{\beta-1} e^{-\frac{\beta h}{\mu}}$$

Comparison Techniques

K-S Test: Compares CDF of proposed PDF to Empirical CDF by calculating the maximum difference.



RMS Test: Compares CDF of proposed PDF to Empirical CDF by summing the square of the differences.

Hellinger Distance: Compares PDF *g(x)* to PDF *f(x)* directly. Yields value of 1 if two PDFs are identical, 0 if two PDFs are disjoint.

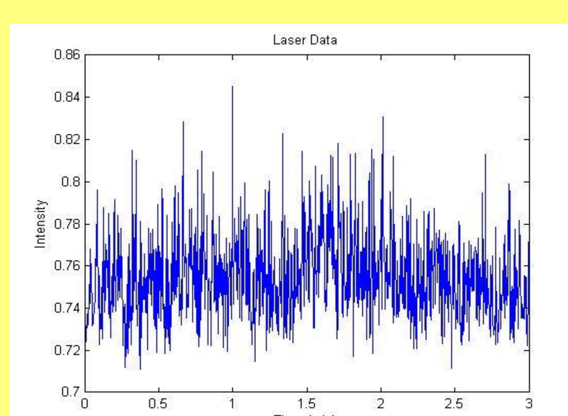
$$k = \sup\{ECDF(x_i) - CDF(x_i):\ i \in N\}$$

$$r = \sum_{i=1}^{N}\left(ECDF(x_i) - CDF(x_i)\right)^2$$

$$h = \int_{-\infty}^{\infty} f(x)g(x)dx$$

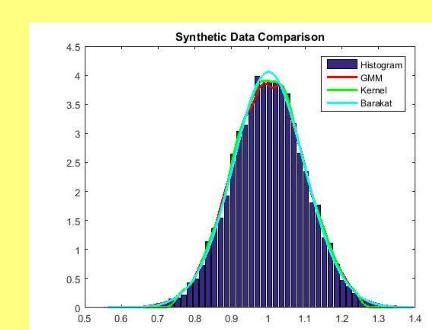
Laser Data Collection

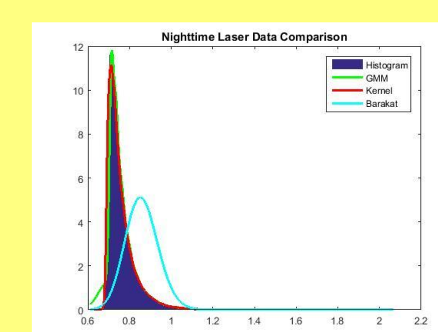
A HeNe laser was propagated through a turbulent atmosphere. The beam was of 632 nm wavelength. The laser was placed in a stationary location and projected a beam 375 meters onto a stationary sensor that recorded intensities of the laser light at a frequency of approximately 10kHz. Recorded for approximately three minutes, this resulted in over 1 million data points in a time series. We have over one hundred data sets on file, taken under different atmospheric conditions, that are usable in MatLab. We investigate the properties of the laser light through these data sets.

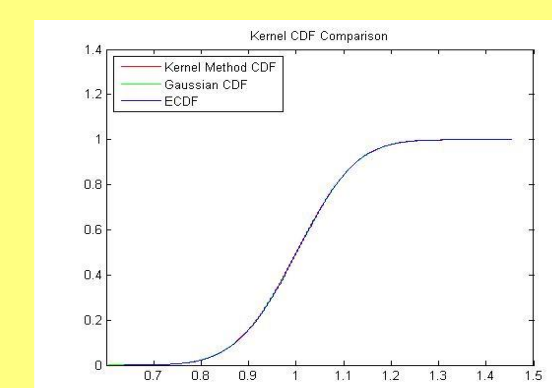


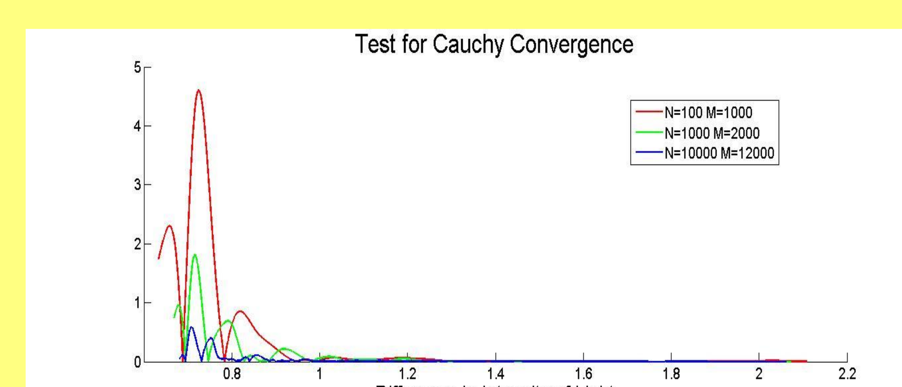
Results

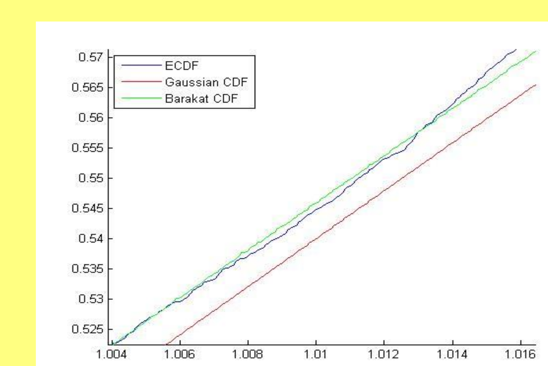
In conclusion, we can see that while the Barakat Method is capable of modeling the synthetic data remarkably well, it cannot account for the noise in the real data. In addition, the beta values from the Barakat Method that result from real data sets are often too large to support numerical approximations computationally. The Kernel Method is capable of representing the empirical data well, but does not yield as much information about the underlying distribution from which the data is pulled (as discovered from the synthetic data simulation). In addition, the Kernel Method is computationally very strenuous. Ultimately, the Gaussian Mixture Method represented the data set well, and was not as computationally strenuous. However, the resulting pdf is not a unique solution, and also requires the user to input the number of clusters beforehand.











Barakat	GMM	Kernel
1.2463358E-04	4.4506809E-04	3.4055791E-03