E205 Lab 1: Data Manipulation and Modeling

Daphne Poon and Sabrina Shen

***Abstract*—This lab involves manipulating and modeling range data from a Velodyne Puck VLP-16 LiDAR sensor, GPS data, and police killings data. The LiDAR data is used to create a probability model and later used to implement Bayes' Rule. All data was provided by the E205 teaching team, and all analysis is completed using Python 3 in a Jupyter Notebook.**

I. Introduction

The provided LiDAR data was collected with a Velodyne Puck VLP-16 LiDAR sensor placed in the Parsons building courtyard. The data files contained Range (m), Latitude and Longitude data, as well as other fields that were not used.

The data is provided for three azimuth values ($-90°, 0°$ and $90°$); for all data points, the elevation angle was $0°$. The range data was obtained in the Parsons building courtyard, and the LiDAR was set up 11 meters from walls North (azimuth = $0°$), West (azimuth = $-90°$) and East (azimuth = $90°$) of the LiDAR.

II. Histograms for Lidar Range Data

A range data histogram was created for each azimuth value (Figs. 1, 2 and 3). We divided the range spread into 50 bins for each histogram.

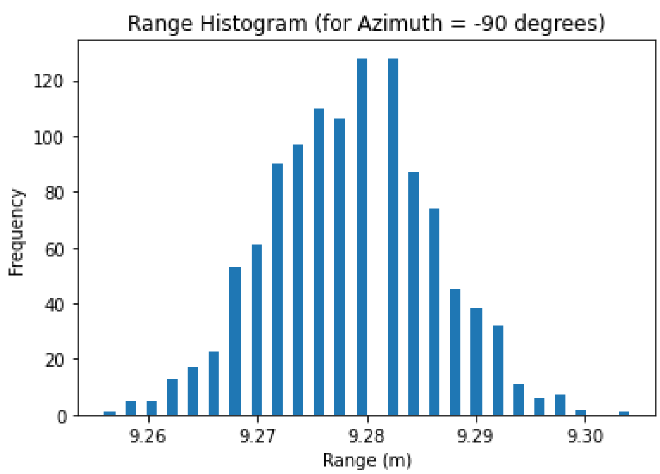

Fig. 1. Range data histogram for azimuth = $-90°$.

III. Create a Model

The beam model of range finders studied in class uses four components to represent the various ways a measurement may be interfered with [1]. For simplicity, we chose to use a Gaussian distribution to model our conditional probability function, with the understanding that within a narrow range of the "true" range z_t, the LiDAR's range distribution looks like a Gaussian distribution.

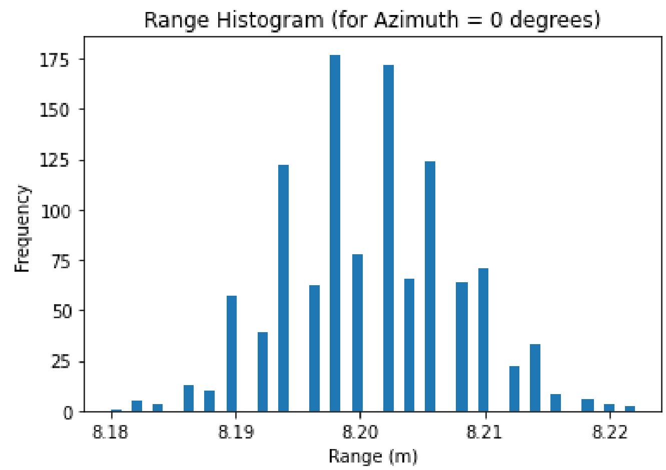

Fig. 2. Range data histogram for azimuth = $0°$.

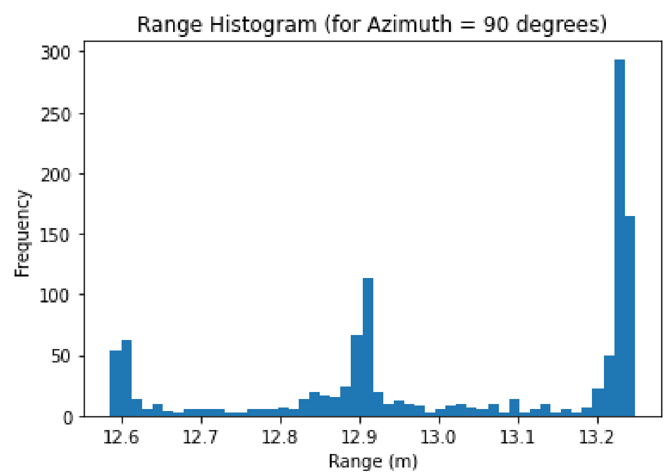

Fig. 3. Range data histogram for azimuth = $90°$.

The conditional probability function we came up with looks as follows:

$$p(z|x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1}$$

where $\mu = 9.281$ and $\sigma = .008$.

Fig. 4 shows how the conditional probability function compares with the original histogram (Fig. 1).

IV. Transform and Plot the GPS Measurements

To transform GPS latitude and longitude data into X-Y coordinates, we used a modified version of the Forward Equirectangular Projection, as described in the E80 lab manual [2].

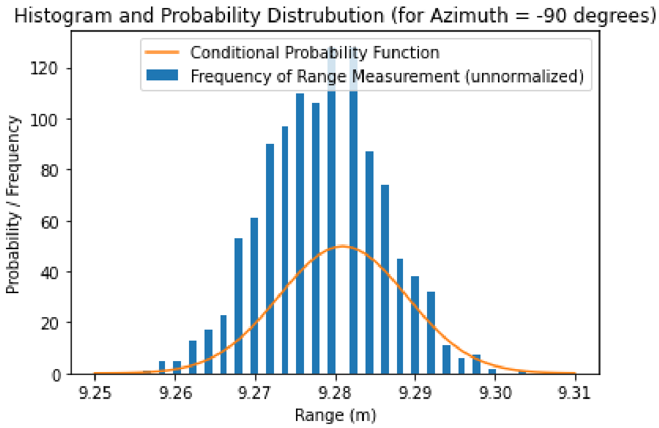

Fig. 4. The conditional probability model is superimposed over the unnormalized histogram for range measurements taken at azimuth $= 90°$.

The equations used to convert latitude and longitude to X-Y are given below as (2), where lat_{curr} and lon_{curr} are the current measurements for latitude and longitude, lat_{origin} and lon_{origin} are defined to be the mean values for latitude and longitude (and used as the X-Y origin), and $R_{Earth} = 6371$ km is the radius of curvature of the Earth.

$$\begin{aligned} y &= R_{Earth} * (lat_{curr} - lat_{origin}) \\ x &= R_{Earth} * (lon_{curr} - lon_{origin}) * \cos(lon_{origin}) \end{aligned} \tag{2}$$

Fig. 5 shows what the X-Y coordinates look like when plotted. We can see that there is much greater variance in the data in the X direction than in the Y direction; the X position is spread over ± 4 m in each direction, while the Y position only ranges over ± 2 m in each direction.

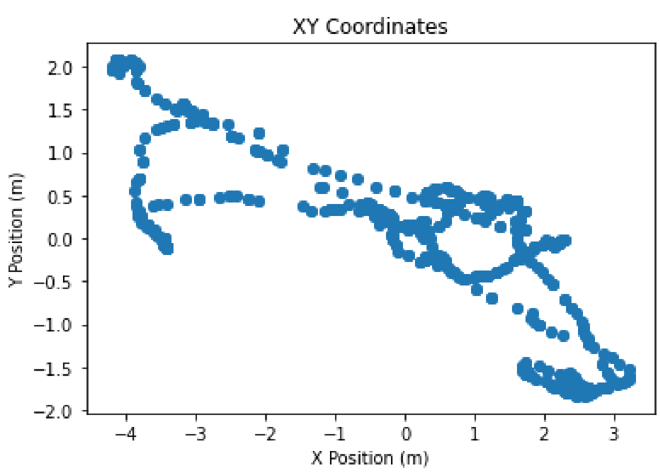

Fig. 5. GPS coordinates converted to XY position using a modified version of the Forward Equirectangular Projection

V. Implementing Bayes' Rule

We are given the four following states $\mathbf{x} = (x, y)$ where the LiDAR could possibly be :

$$\mathbf{x_1} = [-1.4 \;\; 2.8] \quad \mathbf{x_2} = [-1.3 \;\; 2.8]$$

$$\mathbf{x_3} = [-1.2 \;\; 2.8] \quad \mathbf{x_4} = [-1.1 \;\; 2.8]$$

We know that the sensor has to be at one of these states, and for each of these four states, $p(\mathbf{x_i}) = 0.25$.

Bayes' rule states that

$$p(\mathbf{x_i}|z) = \frac{p(z|\mathbf{x_i})p(\mathbf{x_i})}{p(z)} \tag{3}$$

We know that the conditional probabilities for the four states sum to 1, i.e.

$$\sum_{i=1}^{4} p(\mathbf{x_i}|z) = 1 \tag{4}$$

Combining our conditional probability model (1) with (4) and (3), we find that $p(\mathbf{x_i}|z) \approx 0$ for all $\mathbf{x_i}$, so (4) rearranges to become

$$\sum_{i=1}^{4} p(\mathbf{x_i}|z) \approx 4p(\mathbf{x_i}|z) = 1 \rightarrow p(\mathbf{x_i}|z) \approx \frac{1}{4} = 0.25 \tag{5}$$

for all $i \in (1, 2, 3, 4)$.

Using the provided 4 position states, we calculated range measurements for 9.6m, 9.7m, 9.8m and 9.9m away from the west wall, while the $-90°$ azimuth range measurements range from $9.26\text{m} \leq z \leq 9.30\text{m}$. The possible states provided are far above this range, and as such, our conditional probability model predicts that $p(\mathbf{x_i}|z) \approx 0$ for all the x_i values. Using the range information for azimuth$= -90°$ provides no useful information.

VI. Create Police Data Model

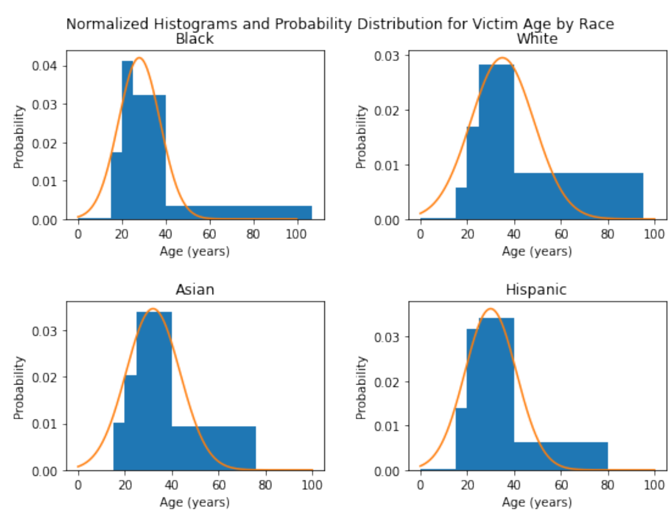

Fig. 6. GPS coordinates converted to XY position using a modified version of the Forward Equirectangular Projection

We created histograms for the age of victims of police killings across various racial backgrounds (Black, White, Asian, Hispanic) using a provided data set. We created probability functions for the following conditional probabilities:

$$p(\text{victims\_age} \mid \text{victims\_race=black})$$
$$p(\text{victims\_age} \mid \text{victims\_race=white})$$
$$p(\text{victims\_age} \mid \text{victims\_race=hispanic})$$
$$p(\text{victims\_age} \mid \text{victims\_race=asian})$$

The data regarding victim's age, separated based on race, was normalized and visualized using the following age bins: 0-15, 16-20, 21-25, 26-40, and 40+. A probability density function was then fit to each distribution as shown in Fig. 6. The mean and standard deviation used to plot each of these functions are listed in Table 1.

TABLE I
μ AND σ FOR PROBABILITY DISTRIBUTION BY RACE

Race	Mean	Standard Deviation
Black	28	9.5
White	35	13.5
Asian	32	11.5
Hispanic	30	11

REFERENCES

[1] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. Probabilistic Robotics (Intelligent Robotics and Autonomous Agents), The MIT Press, 2005.

[2] Lab 7: Navigation, Harvey Mudd College, Claremont, CA, 2018.