Alexander Smith

ECE 3522: STOCHASTIC PROCESSES IN SIGNALS AND SYSTEMS

**COMPUTER ASSIGNMENT (CA) NO. 4: MODEL FITTING**

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122

# Problem Statement

We are given two files: one .xls and one .raw file. The first file’s data set represents Google’s stock price since its inception. The second file’s data is speech sampled at 8 kHz. We are asked to demonstrate how we can model data using a parametric model of a PDF. Our tasks given by our instructor are:

### “ 1) Compute a histogram of the amplitude of the data and normalize it by the number of samples so that it is an estimate of the pdf.

### Fit this distribution by estimating the mean and variance. Plot the Gaussian model on top of the histogram. Compare and contrast the quality of the fits to the data.

### In (2), you should find that the Gaussian model is not a good fit for the Google data. Select another distribution from Chapter 4 that provides a better estimate of the data. Plot this model on the same graph with the histogram and the Gaussian fit. Compute the mean-squared error between the actual data and the parametric fit. Which gives a better approximation? (Do this for both data sets.) “

The first task can easily be done using the *histogram* function with its normalization property set to *‘pdf’*. The distribution line fit can be done using the *fitdist* function and manually plotting the resulting array as a function of values that span those of the sample space.

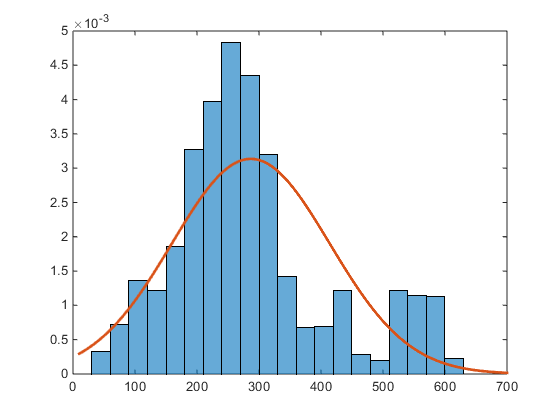


Figure 1. Closing stock histogram PDF estimate with normal/Gaussian distribution overlaid fit

The Gaussian distribution fit is not a good fit for the stock PDF.

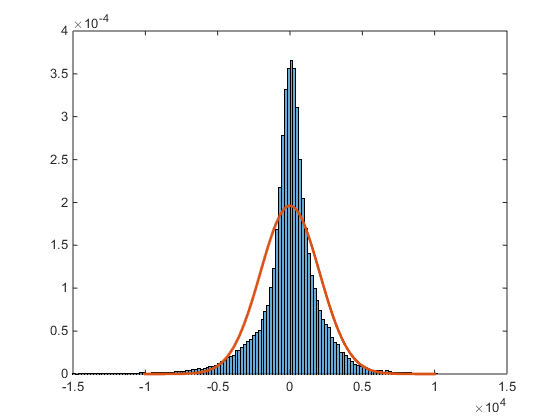


Figure 2. Speech signal histogram PDF estimate with normal/Gaussian distribution overlaid fit.

My Gaussian distribution fit is almost a good fit for the speech histogram.

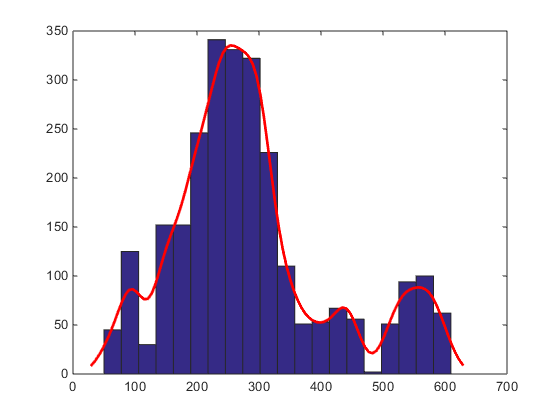


Figure . Speech signal histogram with kernel overlaid fit

I tested most of the different distribution fit choices that MATLAB offers for its *histfit* function, and found ‘*kernel’* to be the best fit for the stock histogram. The ‘*kernel’* choice in MATLAB generates a sample of size 100 from a normal distribution with mean 10 and variance 1 to make the fit.

# MATLAB Code

%

clear; clc; close all;

%Create signal arrays

ggl = xlsread('google\_v00.xlsx'); %Stock

ggl=ggl(:,4); % only want closing stock

%

Faudio = fopen('rec\_01\_speech.raw','r');

spch = fread(Faudio, inf, 'int16'); %Speech

fclose(Faudio);

pd1 = fitdist(ggl,'Normal')

pd2 = fitdist(spch,'Normal')

% pdf histogram

figure

histogram(ggl,'Normalization','pdf')

hold on

x\_values = 10:1:700;

y = pdf(pd1,x\_values);

plot(x\_values,y,'LineWidth',2)

figure

histogram(spch,'Normalization','pdf')

hold on

x\_values = -10000:1:10000;

y = pdf(pd2,x\_values);

plot(x\_values,y,'LineWidth',2)

figure

histfit(ggl,20,'kernel')

%{

figure

histfit(spch,20,'normal')

%}

# Conclusions