Marc Jurchak

ECE 3522: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 1912

# Problem Statement

 We will model the speech signal and the Google stock prices with parametric models of the PDF. We begin by estimating the mean and variance of the signals, and then move on to plotting distributions over the PDF to see which fits the best. To determine the best fit for both signals, we compute the mean-square error and see which distribution has the lowest.

# Approach and Results

We begin by computing the mean and variance of each signal, and then plotting them over the PMF. These results are shown in figures 1 and 2. The stock price is the signal average with frame size 1 and window size 30, and the speech signal is averaged at frame 10 and window 240. This is done for easier viewing. Note all calculations are done with the original data.



Figure 1: stock PMF, average value and variance.



Figure 2: speech PMF, average value and variance.

We obtain the distributions with MATLAB’s “fitdist()” and “pdf()” functions.

 We see that there are distributions that can fit the Google stock prices better, so we try different distributions with the “dfittool”. From the book we find that the Erlang distribution should fit our data nicely. From the book, “the Erlang random variable describes the time interval between any event and the kth following event.” This should give it the dexterity to match our distribution. We find that the Gamma distribution, which is a special case of the Erlang distribution, fits the stock PMF well. We also find that the Normal distribution fits the speech signal the best, but the Burr distribution is close. These results are shown in figures 3 and 4.



Figure 3: Google stock PMFs.



Figure 4: Speech signal PMFs.

 We take the mean-square error to decide which distributions are mathematically better. We also remember that the plots are filtered versions of the signals, and that the real PMF is very noisy. Thus some distributions will be more mathematically accurate even if their plots do not look it. MSEs are in table 1.

|  |  |  |  |
| --- | --- | --- | --- |
|   | Normal | Gamma | Burr |
| Speech | 1.50E+04 |   | 3.00E+04 |
| Stock | 4.67E-06 | 3.56E-19 |   |

Table 1: MSE of the distributions.

# MATLAB Code

clear;

clc;

clf;

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

speech = fread(fp,inf,'int16');

fclose(fp);

google\_stock = xlsread('google\_v00.xlsx');

stock = google\_stock(:,4)';

shifted\_speech = speech + 15000;

% build stock PDF

posedge\_stock = 1:1:2620;

N\_stock = histcounts(stock,posedge\_stock);

stock\_pdf = N\_stock ./ length(stock);

% get average value of PDF so plot is cleaner

[filtered\_stock\_pdf, vari] = compute\_avg\_and\_var\_v01(stock\_pdf, 1, 30);

% build Google PDF

%edge\_speech = min(speech):1:max(speech);

edge\_speech = 1:1:length(shifted\_speech);

N\_speech = histcounts(shifted\_speech,edge\_speech);

speech\_pdf = N\_speech ./ length(shifted\_speech);

% get average value of PDF so plot is cleaner

[filtered\_speech\_pdf, vari] = compute\_avg\_and\_var\_v01(speech\_pdf, 10, 240);

% construct distributions from both data.

stock\_gamma = fitdist(stock','gamma');

stock\_gamma\_pdf = pdf(stock\_gamma,(1:1:2619));

stock\_normal = fitdist(stock','normal');

stock\_normal\_pdf = pdf(stock\_normal,(1:1:2619));

speech\_normal = fitdist(shifted\_speech,'normal');

speech\_normal\_pdf = pdf(speech\_normal,(1:1:length(shifted\_speech)));

speech\_burr = fitdist(shifted\_speech,'Burr');%tLocationScale

speech\_burr\_pdf = pdf(speech\_burr,(1:1:length(shifted\_speech)));

figure(1)

hold on

plot(stock\_gamma\_pdf,'b', 'DisplayName','gamma distribution')

plot(stock\_normal\_pdf,'r', 'DisplayName','normal distribution')

plot(filtered\_stock\_pdf,'k', 'DisplayName','stock PDF')

plot([1 length(stock\_pdf)],[mean(stock\_pdf) mean(stock\_pdf)],'r-.','DisplayName','mean')

plot([1 length(stock\_pdf)],[var(stock\_pdf) var(stock\_pdf)],'b-.','DisplayName','variance')

axis([0 1000 0 6e-3])

ylabel('probability')

xlabel('stock price')

title('Google stock PMF')

legend('-DynamicLegend')

hold off

figure(2)

hold on

plot(speech\_normal\_pdf,'r','DisplayName','normal distribution')

plot(filtered\_speech\_pdf,'k','DisplayName','speech PDF')

plot(speech\_burr\_pdf,'b','DisplayName','Burr distribution')

plot([1 length(speech\_pdf)],[mean(speech\_pdf) mean(speech\_pdf)],'r-.','DisplayName','mean')

plot([1 length(speech\_pdf)],[var(speech\_pdf) var(speech\_pdf)],'b-.','DisplayName','variance')

axis([0 3e4 0 4e-4])

ylabel('probability')

xlabel('time')

title('Speech signal PMF')

legend('-DynamicLegend')

hold off

% Calculate mean-squared error for gamma distribution for Google stock

sum = 0;

for i = 1:length(stock\_pdf)

 sum = sum + (stock\_pdf(i) - stock\_gamma\_pdf(i));

end

MSE\_stock\_gamma = sum/length(stock\_pdf)

% Calculate mean-squared error for normal distribution for Google stock

sum = 0;

for i = 1:length(stock\_pdf)

 sum = sum + (stock\_pdf(i) - stock\_normal\_pdf(i));

end

MSE\_stock\_normal = sum/length(stock\_pdf)

% Calculate mean-squared error for normal distribution for speech signal

for i = 1:length(shifted\_speech)

 sum = sum + (shifted\_speech(i) - speech\_normal\_pdf(i));

end

MSE\_speech\_normal = sum/length(shifted\_speech)

% Calculate mean-squared error for burr distribution for speech signal

for i = 1:length(shifted\_speech)

 sum = sum + (shifted\_speech(i) - speech\_burr\_pdf(i));

end

MSE\_speech\_burr = sum/length(shifted\_speech)

# Conclusions

 We find that it was difficult to match a distribution to the PMFs, since the unfiltered PMF is so noisy it is difficult to see if the distribution matches it, but the filtered versions, as above, do not quite match the error suggested in table 1. For example the Google stock Normal distribution seems only a little offset from the Gamma, but the mean-square error (MSE) suggests that the Gamma distribution is many order of magnitudes better. Also, the Burr distribution seems to better represent the PMF, but the normal distribution has a lower MSE. We also found that the average value of the PMF and the variance do not give much information, since many data points have a 0 probability. This skews the constant average with information we are not concerned with.