Data Analysis and Interpretation

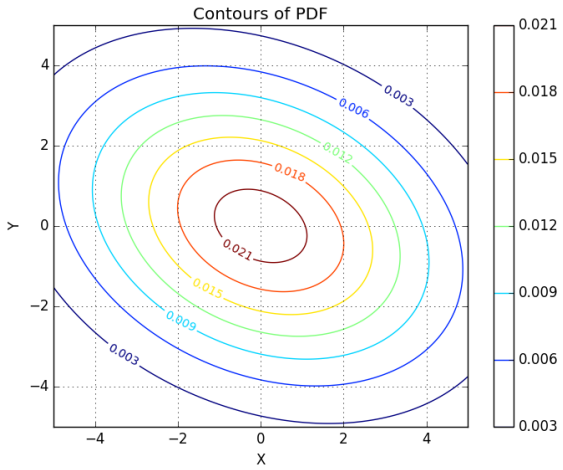
Assignment Report 2

Rahul Dandwate
Guru Vamsi
Amey Gaikwad
Sumukh Vaidya

October 13, 2016

1. Question 1

1.1 Plotting contour for the Distribution

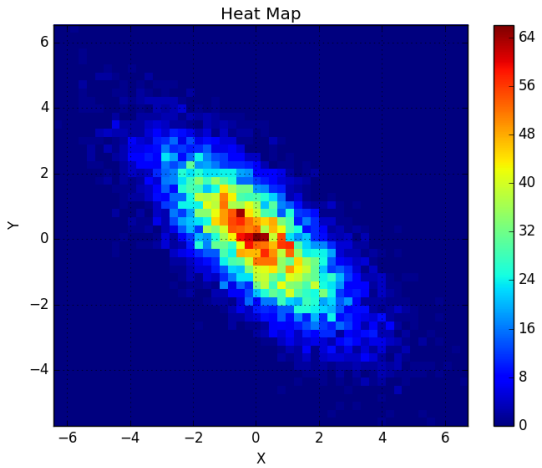


The contours is plotted using the contour function in Matplotlib.

1.2 Generating random variables

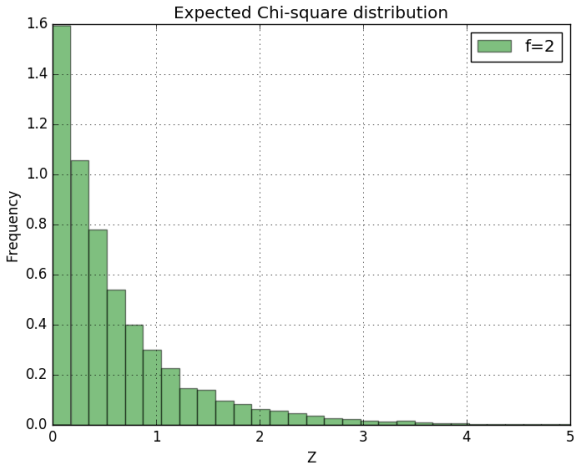
A pair of random variables (x, y) are generated provided the covariance matrix using Multivariate Normal function of Numpy.

1.3 Heat Map



The Heat Map produced provided the covariance matrix using hist2d function of Matplotlib.

1.4 Histogram of random variable z



Mean value of z = 2
Standard Deviation of z = 2
Yes, this is consistent with the expectation as after making the given substitution the distribution becomes a $\chi^2_2(z)$ distribution.
$\chi^2_2(z) = \frac{e^{\frac{-z}{2}}}{2}$

2. Question 2

Mean Weight of the Panda $(\overline{w}) = \sum_{i=1}^{N} \frac{w_i}{N}$ = 49.41660999999994 kg

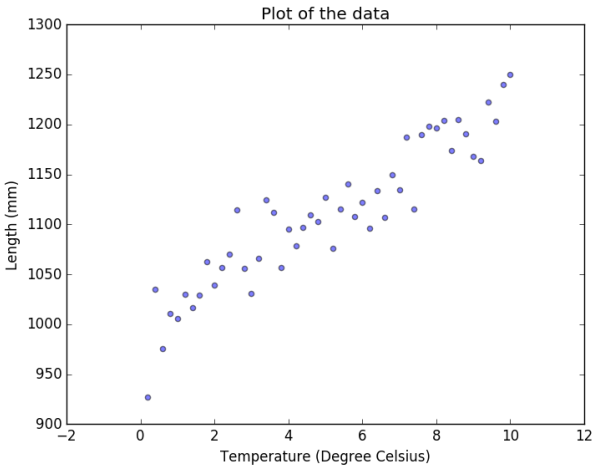
Error in mean weight of Panda= $\sqrt{\sum_{i=1}^{N} \frac{(w_i - \overline{w})^2}{(N-1)(N)}}$ = 0.322898739861 kg

Error in weight of individual Panda = $\sqrt{\sum_{i=1}^{N} \frac{(w_i - \overline{w})^2}{(N-1)}}$ = 10.2109547156 kg

3. Question 3

3.1 Plotting given Data

The given was plotting using the scatter plot function in Matplotlib in python.
Following is the plot of the data:

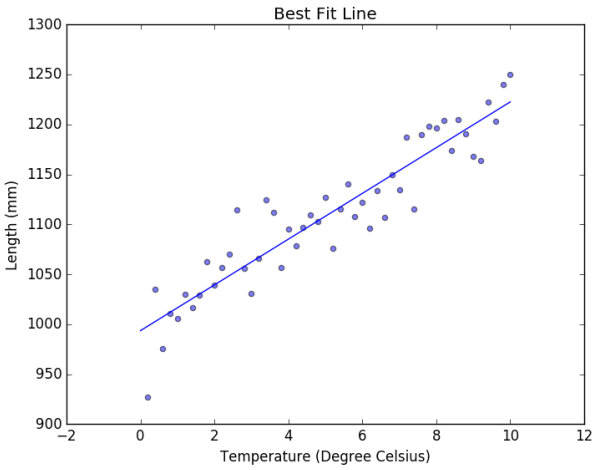


3.2 Minimizing the least square distance

For finding the best fit line through the data the Least Square Distance is employed.
Least Square Distance = $\sum_{i=1}^{N} \frac{((a \times x_i) + b - y_i)^2}{(N-2)}$
where a and b are the parameters of the best fit line.
N = 50
The parameters are determined using Gradient Descent Algorithm.
Expected Length at 0 Degree C (a) = 993.513738775 mm
Coefficient of linear expansion (b) = 22.9018943577 $\frac{mm}{Kelvin}$

3.3 Extrapolation of the curve

Expected Length at 15 Degree C = $(15 \times 22.9018943577) + 993.513738775 = 1337.04215414$ mm
Error in Length at 15 Degree C = 26.7137818002 mm
The error in the length is the square root of the minimized value of the above Least Square Distance Function for all the 50 values of the data.



4. Distribution of work for each week

The work distribution for the week was as follows:

Team Role Assignment		
Week No.	Roles Assigned	Names
Week 1	Team Leader	Amey Gaikwad
	Programmer	Rahul Dandwate, Guru Vamsi, Amey Gaikwad
	Report Writer	Sumukh Vaidya
	Website Manager	Sumukh Vaidya
Week 2	Team Leader	Rahul Dandwate
	Programmer	Rahul Dandwate, Guru Vamsi, Amey Gaikwad
	Report Writer	Rahul Dandwate
	Website Manager	Sumukh Vaidya