

---

# Statistical Methods for Machine Learning

## *First Assignment*

---

by:

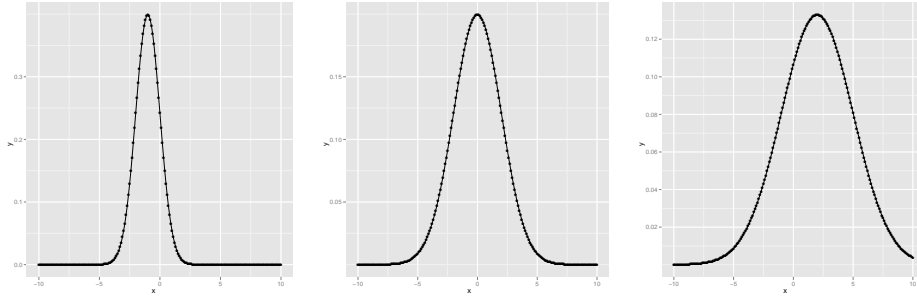
*Nicholas Clark, Piotr Zywien, Victor Yakimov*

## README

All the code has been run and tested in R 2.13.1 using RStudio 0.94.110 on an Apple x86\_64, darwin9.8.0 system. Packages required are `pixmap`, `ggplot2`, and `gplots`. A current version of R can be obtained directly from CRAN's website, <http://cran.r-project.org>, and packages can be installed from the CRAN repository by running the command `install.packages` from the R environment.

### QUESTION 1

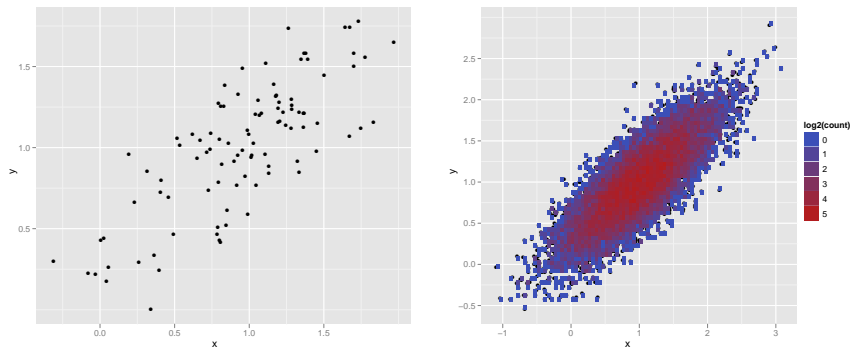
The normal distribution with parameters  $\mu, \sigma = (-1, 1), (0, 2), (2, 3)$  can be seen in figure 1.



**Figure 1.** Gaussian distributions with  $\mu, \sigma = (-1, 1), (0, 2), (2, 3)$

### QUESTION 2

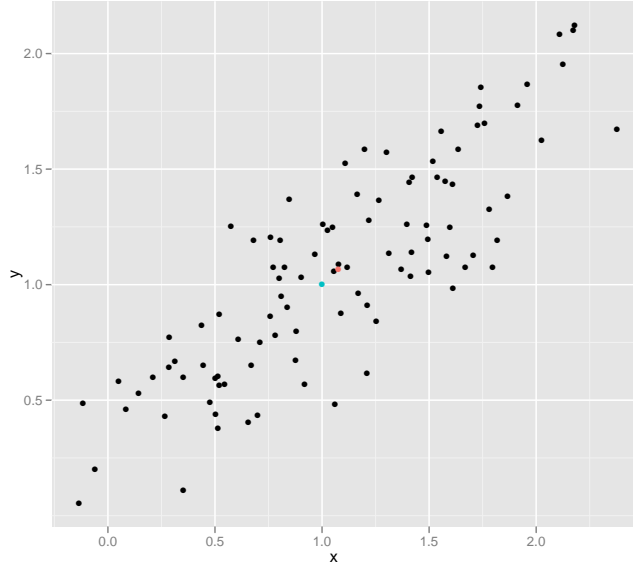
In figure 2 we have two empirical plots of a 2-dimensional gaussian, with  $\mu = (1, 1)^T$  and  $\sigma = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}$ . One has a sample size  $N = 100$ , and the other a sample size  $N = 10000$ , with binned counts shown as a heat-map.



**Figure 2.** One hundred samples drawn from a 2-dimensional gaussian, and one hundred thousand samples drawn from a 2-dimensional gaussian with binned counts, drawn for comparison

### QUESTION 3

The sample mean and covariance is  $\mu_{ML} = (1.076, 1.065)^T$  and  $\Sigma_{ML} = \begin{bmatrix} 0.340 & 0.225 \\ 0.225 & 0.210 \end{bmatrix}$ . The sample mean and the correct mean can be seen on figure 3. The two differ, as the samples drawn from a gaussian are naturally stochastic. The sample mean and the true mean will converge as the sample number  $N \rightarrow \infty$ . In our case the sample mean deviates by  $(0.076, 0.065)$  from the true mean



**Figure 3.** One hundred samples drawn from a 2d-gaussian distribution. The blue dot represents the true mean of the distribution used, and the red dot represents the sample mean

### QUESTION 4

In order to show an expression of conditional trivariate Gaussian distribution,  $p(x_a|x_b)$  with  $x_c$  marginalized out, we use formula (2.97) from Bishop:

$$\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab}(x_b - \mu_b)$$

and equations (2.79)

$$\Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$$

and (2.80)

$$\Lambda_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$$

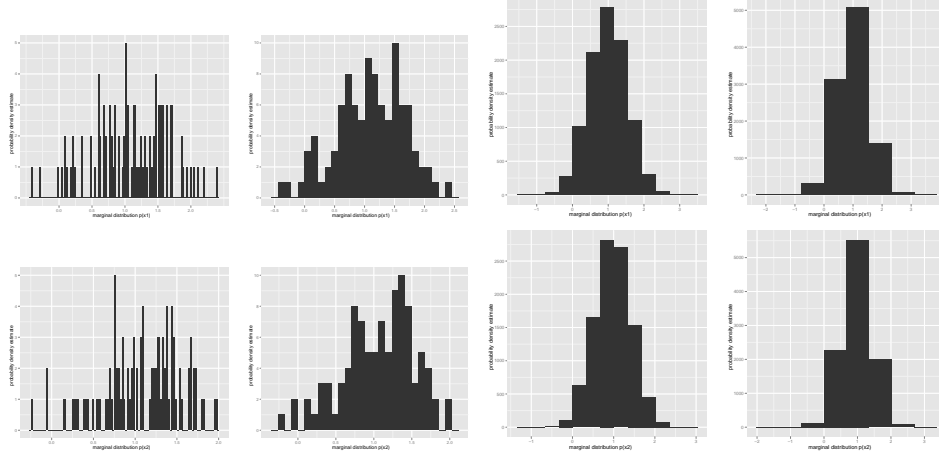
Insertion gives

$$\begin{aligned} \mu_{a|b} &= \mu_a - (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})(-(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1})(x_b - \mu_b) \\ &\Leftrightarrow \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) \end{aligned}$$

And by equation 2.79,  $\Lambda_{aa}^{-1} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ . We can insert those values into  $p(x_a|x_b) = \mathcal{N}(x|\mu_{a|b}, \Lambda_{aa}^{-1})$

### QUESTION 5

As is apparent from figure 4, large bins result in a loss of detail (as is the case with 5 bins in the figure), while small bins can often be misleading as they are too sensitive to outliers (as is the case with 100 bins).



**Figure 4.** The first dimension  $p(x_1)$  (top row) and second dimension  $p(x_2)$  (bottom row) of one hundred samples from a 2-dimensional gaussian represented as a histogram, with different bin widths. The number of bins is 100, 25, 10 and 5, per column

### QUESTION 6

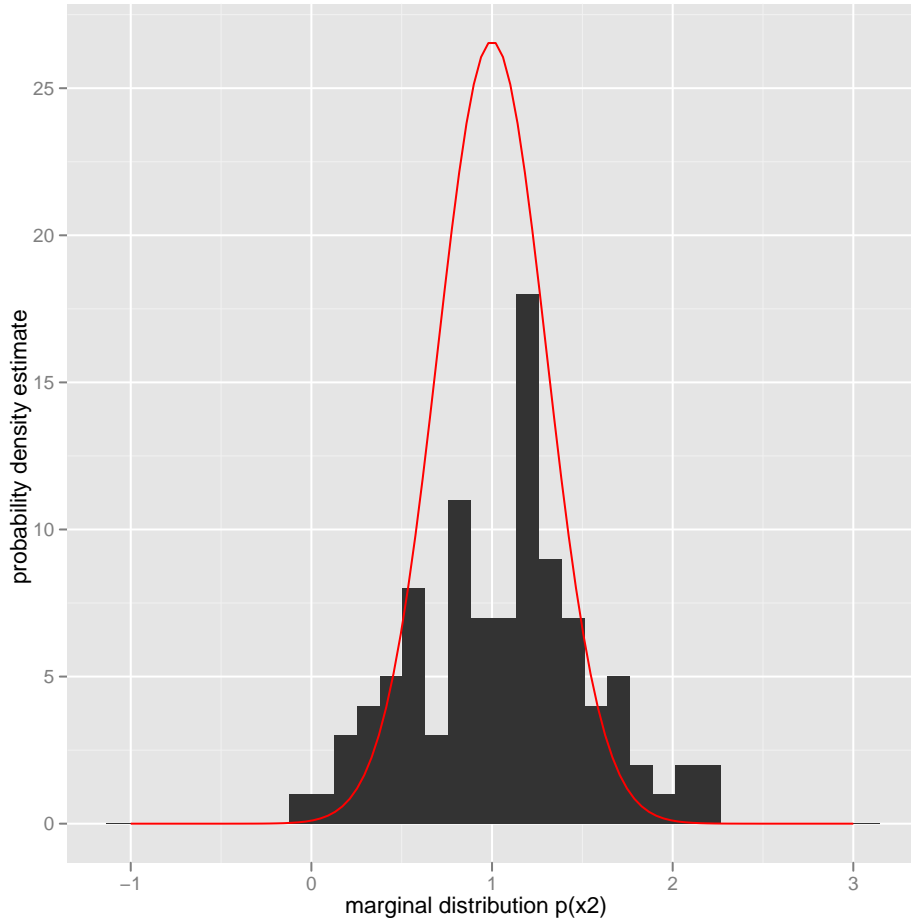
The histogram for the first dimension  $p(x_1)$  of one hundred samples from a 2-dimensional gaussian, plotted along with its analytical solution, with  $\mu = 1$ ,  $\sigma = 0.3$  shown in figure 5. The analytical expression for the marginal distribution  $p(x_1)$  is

$$\mathcal{N}(\mu = \mu_1, \sigma^2 = \sigma_{11}^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

given a bivariate Gaussian with  $\mu = (\mu_1, \mu_2)^T$  and  $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

### QUESTION 7

In figure 6 we show 2-dimensional histogram estimate of  $p(x)$ , using 100,1000,10000 samples and  $10 \times 10$ ,  $15 \times 15$ , and  $20 \times 20$  bins. Larger number of samples mean we can decrease the bin-width further and provide more resolution of the histogram. Histograms with small bin sizes and few samples tend to behave more erratically – i.e. if we were to replot all the histograms, the one with the least samples and smallest bins would vary most.

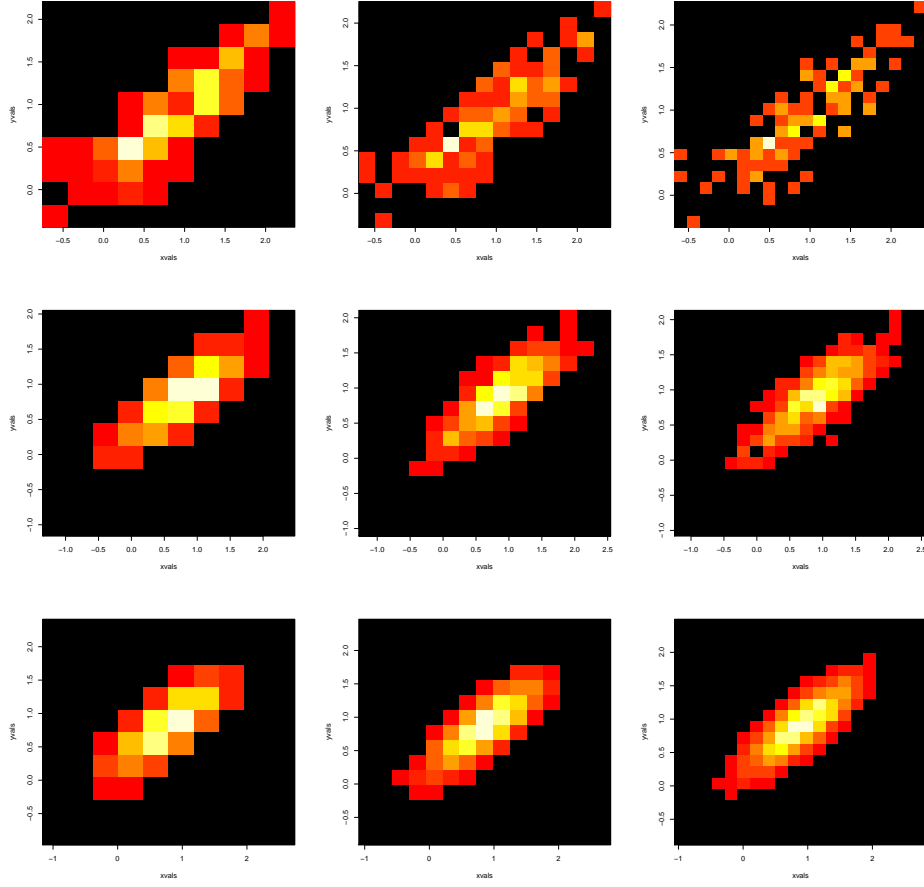


**Figure 5.** The first dimension  $p(x_1)$  of one hundred samples from a 2-dimensional gaussian represented as a histogram, plotted along with the analytical solution for the mean and covariance parameters.

#### QUESTION 8

We derive a function  $h(z) = 1/\lambda \cdot \ln(1 - z)$  which is the integral of  $p(y)$ . We then generate a uniformly random number  $z \in [0, 1]$   $L=10,100,1000$  times which we use to estimate  $\mu$  of  $p(y)$ . Repeating the estimation 1000 times, we get a gaussian distribution of means, as predicted by the central limit theorem. A normal quantile-quantile plot, as well as a histogram of the means, shown in figure 7 shows that a thousand means drawn from  $L=1000,100,10$  very closely approximate a Gaussian distribution, centered around the true mean of  $\mu = 1/\lambda = 1/3$ . We get a standard deviation of the mean of  $\sigma = 0.104$ ,  $\sigma = 0.0324$ , and  $\sigma = 0.0108$  for  $L=1000$ ,  $L=100$ , and  $L=10$ , respectively.

We also show in figure 8 how the sample mean deviates from the true mean with varying values of  $L$  sizes. In general, for all three values of  $L$ ,



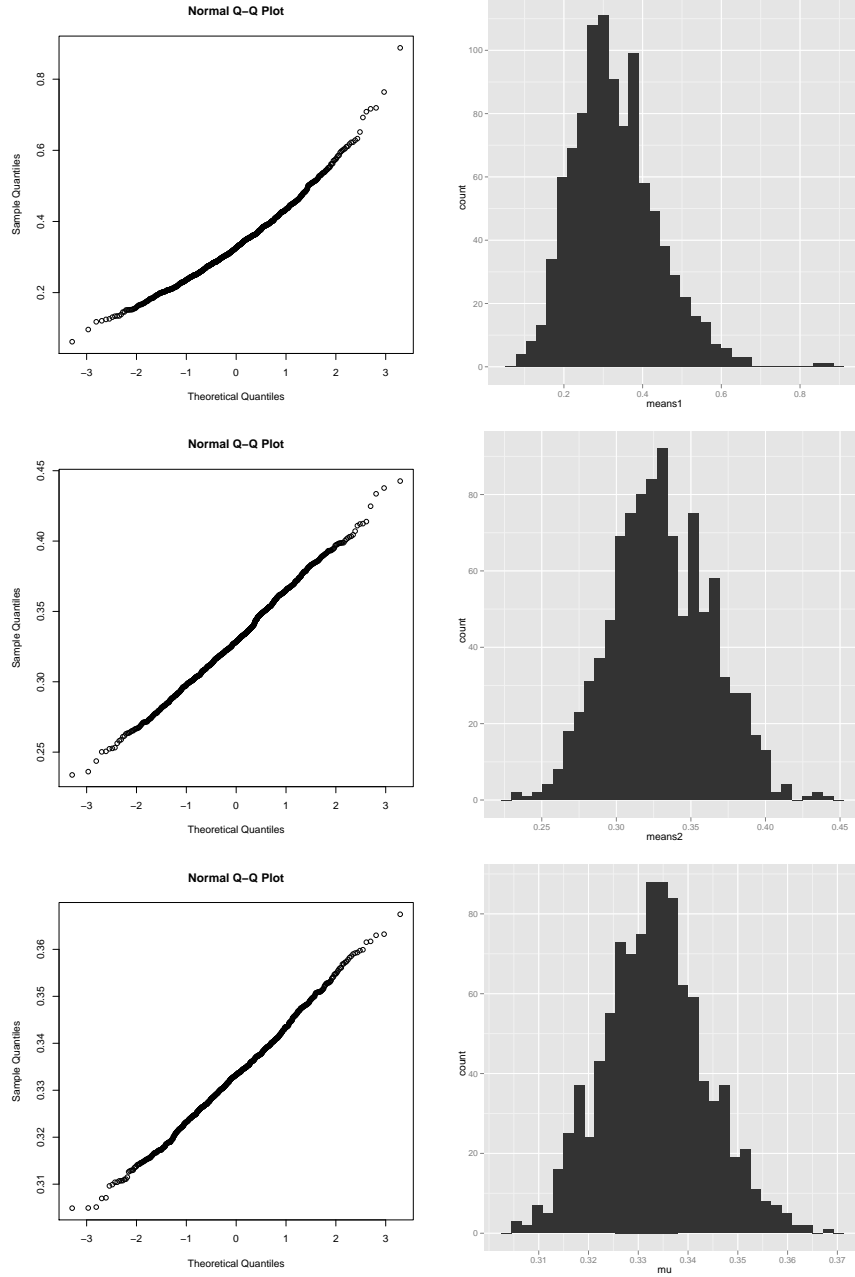
**Figure 6.** 2D-histograms. First row contains 100 samples, second row contains 1000 samples, and the third row contains 10000 samples. The first column has  $10 \times 10$  bins, the second  $15 \times 15$  bins, and the third  $20 \times 20$  bins

the sample mean should be around the true mean, with the only difference being the variance of means.

A transformation of the expected absolute deviation, when plotted against  $L$  will produce a straight line, given the transformation is the inverse of the square of said mean absolute deviation. Thus, the plotted points will be as such:  $(10, \frac{1}{\text{mad}(\text{means1})^2})$ ,  $(100, \frac{1}{\text{mad}(\text{means2})^2})$  and  $(1000, \frac{1}{\text{mad}(\text{means3})^2})$ , where  $\text{mad}$  is the mean absolute deviation function and  $\text{means1}$ ,  $\text{means2}$  and  $\text{means3}$  are mean values plotted in figure 7. The straight plotted line is shown in figure 9. This reflects the property that in order to achieve higher precision equivalent to the square root of the standard deviation, the amount of data has to be proportionally increased by a factor of 10.

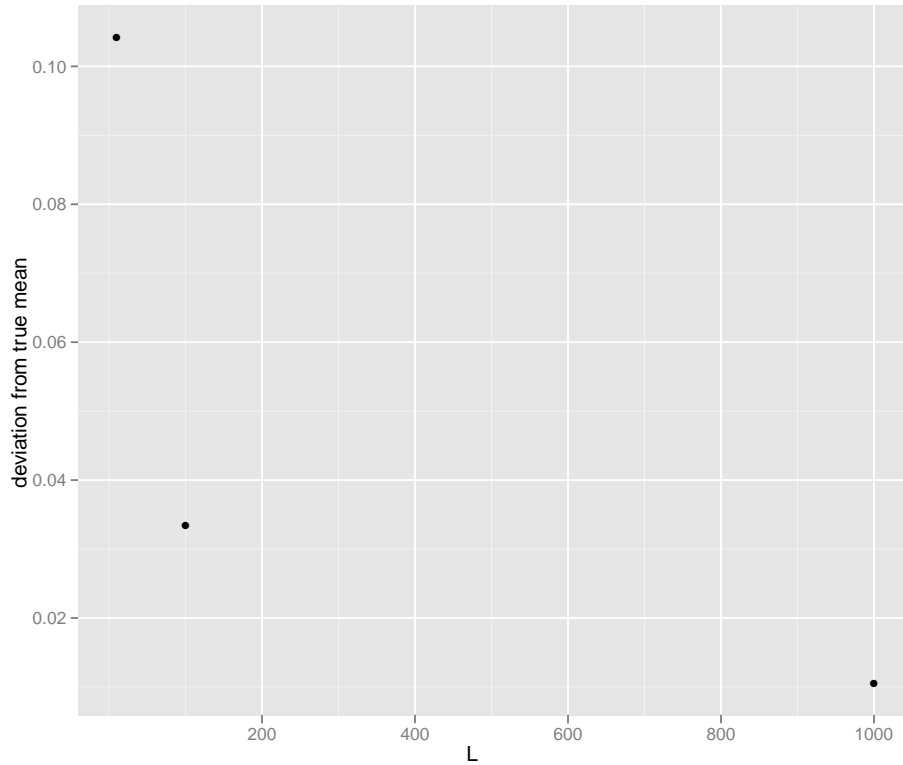
## QUESTION 9

The red pitcher is detected with remarkable accuracy, and the probability density is shown in figure 10. The pitcher is detected, and no pixels from



**Figure 7.** Quantile-quantile plot and histogram for the distribution of  $\mu$  with  $L=10,100,1000$ , shown in rows 1, 2, and 3

outside the pitcher are picked up. Specular highlights present in the center of the image are not detected as they do not have the same color distribution as the training set — the training set is mostly red, and the specular highlights are mostly white.



**Figure 8.** The deviation from the true mean with increasing  $L$

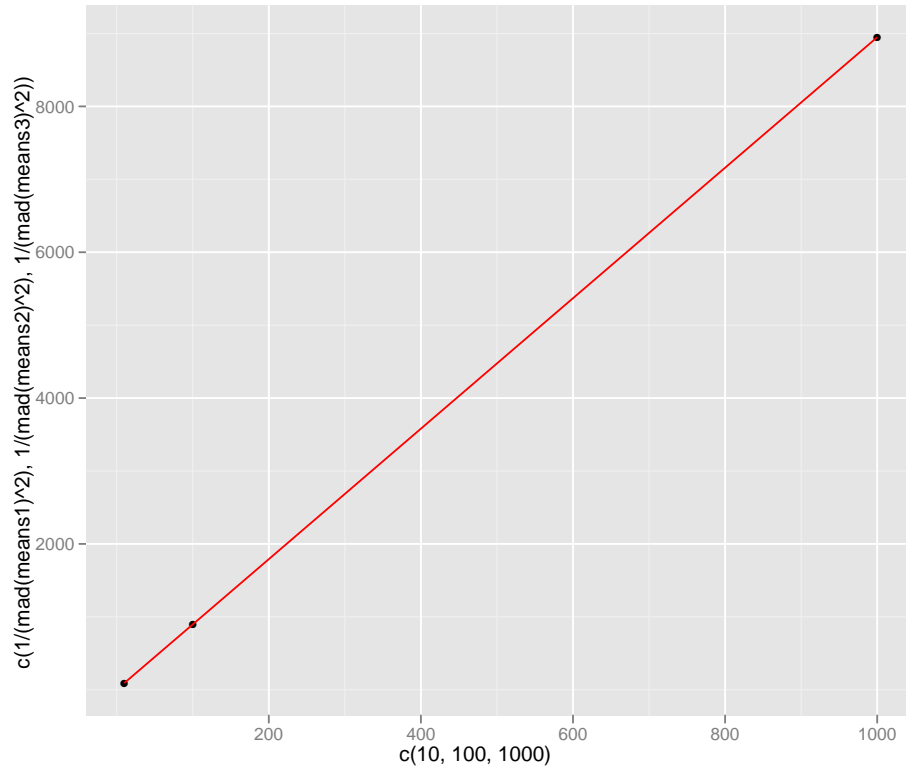
#### QUESTION 10

The center of mass can be seen in figure 10 and the iso-probability curves along with the spatial covariance are shown in figure 11.

#### QUESTION 11

We test our trained model on a different picture of a pitcher. We use the same training data, and display the probability that pixels from the new image are a part of the old distribution. We get the 2-d probability map shown in figure 11. The results are poor, which is unsurprising as the pitcher appears as a significantly different shade of red under the different lighting conditions.

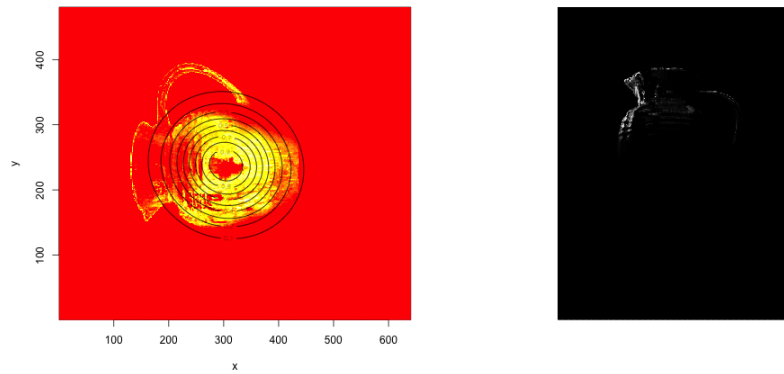




**Figure 9.** Plot of the inverse squares marginal average distributions of means against L, where L is 10, 100 and 1000



**Figure 10.** probability density of pitcher detection, along with its center of mass



**Figure 11.** iso-probability curves and 2-d probabilities for kande2.png