Math 537 HW 2

Gustavo Esparza

7/12/2019

1A

The unbiased estimator of Σ, $S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})'(X_i - \bar{X})$:

	x1	x2
x1	5.63	3.16
x2	3.16	12.18

1B

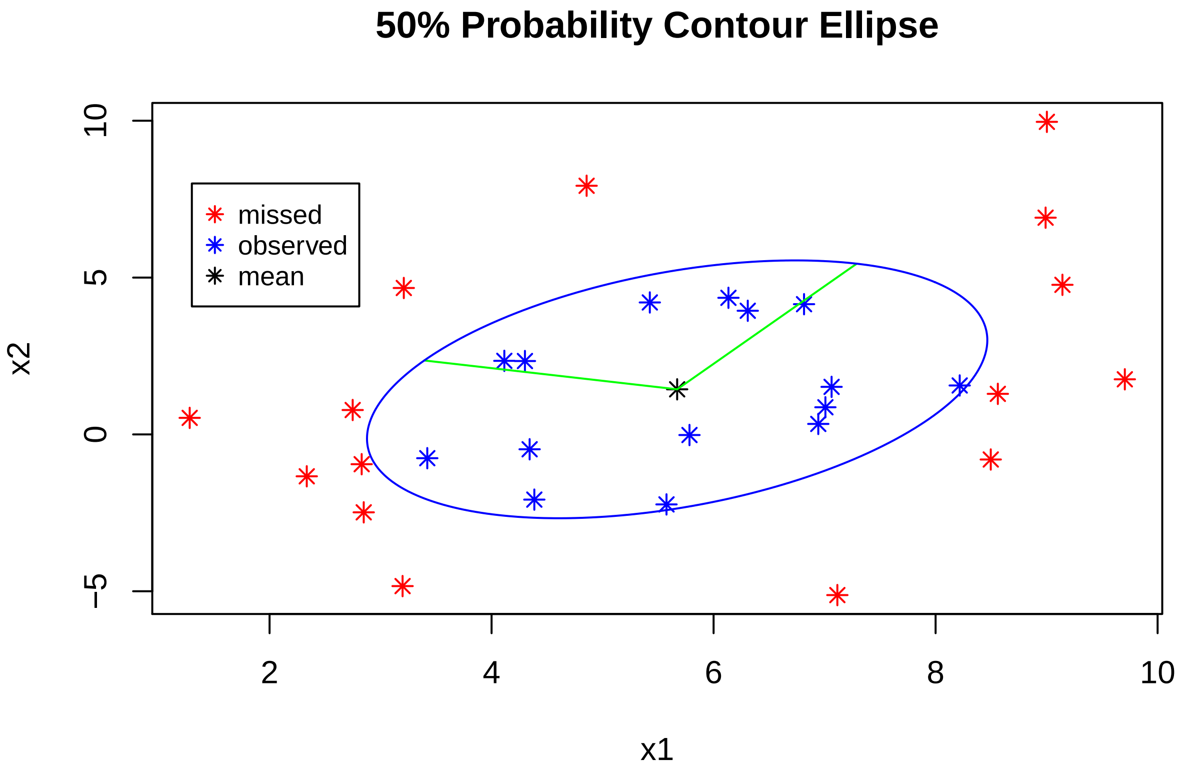
For B, we will use the following formula to compute the distances : $d^2 = (\vec{X} - \bar{X})'\Sigma^{-1}(\vec{X} - \bar{X})$
Here is a table counting our original (x_1, x_2) observations, along with their respective distance from $\bar{x} = (5.671873, 1.437181)$.

	x1	x2	Distance
1	4.11	2.35	0.74
2	3.21	4.67	3.12
3	3.20	-4.84	3.38
4	9.14	4.77	2.32
5	5.43	4.21	0.82
6	3.42	-0.76	0.98
7	8.56	1.29	1.78
8	7.01	0.86	0.48
9	6.81	4.15	0.64
10	7.11	-5.12	5.59
11	8.99	6.91	3.21
12	6.31	3.94	0.51
13	2.85	-2.48	1.94
14	4.86	7.93	4.76
15	6.94	0.33	0.60
16	4.39	-2.08	1.04
17	5.78	-0.02	0.22
18	7.06	1.52	0.39
19	4.30	2.34	0.60
20	9.00	9.96	6.23
21	4.34	-0.47	0.44
22	6.13	4.36	0.72
23	2.34	-1.34	2.05
24	5.58	-2.23	1.26
25	9.70	1.76	3.25
26	8.50	-0.80	2.82
27	2.75	0.78	1.61
28	8.22	1.56	1.31
29	2.83	-0.95	1.49
30	1.28	0.53	3.66

1C

```
## The Chi-Square value corresponding to a 50% probability contour =  1.386294
## The number of observations that fall within this distance       =  15
## The proportion of observations that fall within the proability  =  0.5
```

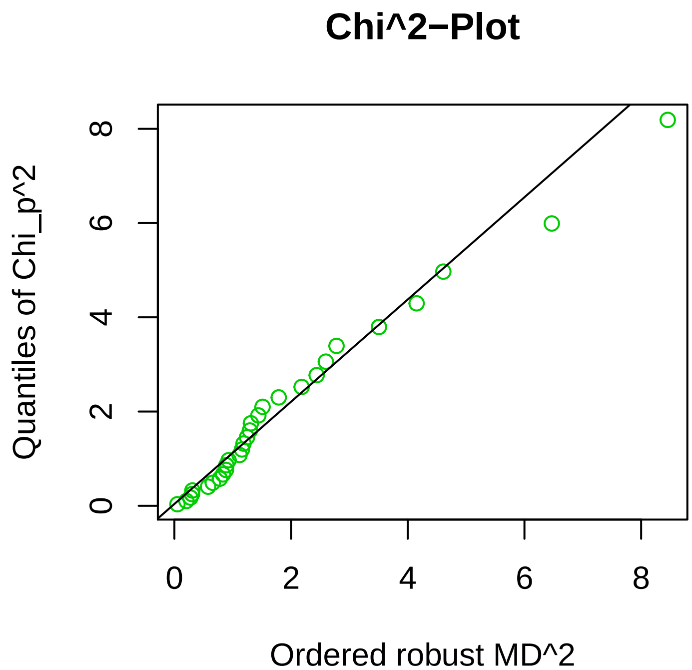
Thus, we can observe that fifty percent of the observed values lie within the fifty percent probability contour. We can visualize this contour probability with an ellipse:



1D

Based on the ellipse showing fifty percent of our data falling within the fifty percent contour, I believe this data to reflect a bivariate normal density. We will plot a $\chi^2 QQ$ plot to emphasize this normality.

Here is our χ^2 plot:



We can see that the plotted c^2 distances have the desired diagonal line that indicates normality for our bivariate data. There is a slight deviation at the tail end, but it is not significant enough to suggest non-normality.

2

For this problem, we have $\vec{X_1}, \vec{X_2}, ..., \vec{X_{20}}$ as observations from a $N_6(\mu, \Sigma)$ distribution.

A

For $(X_1 - \mu)'\Sigma^{-1}(X_1 - \mu)$, we will use result 4.7 which states that this product follows a χ_p^2 distribution. In this instance, we have $p = 6$, thus $(X_1 - \mu)'\Sigma^{-1}(X_1 - \mu) \sim \chi_6^2$

B

If we $X \sim N_p(\mu, \Sigma)$, then we know that $\bar{X} \sim N_p(\mu, \frac{\Sigma}{n})$.
Then we have the following result: $\bar{X} \sim N_6(\mu, \frac{\Sigma}{20})$

C

Given our result in B, we can see that $\bar{X} - \mu \sim N_6(\mu - \mu, \frac{\Sigma}{20}) = N_6(0, \frac{\Sigma}{20})$. This is a result in subtracting μ affecting the mean alone.
Now, we have $\sqrt{n}(\bar{x} - \mu)$ being a multiplier, and therefore affecting our variance. In particular, we have : $\sqrt{n}(\bar{x} - \mu) \sim \sqrt{20}N_6(0, \frac{\Sigma}{20}) = N_6(0, 20 \times \frac{\Sigma}{20}) = N_6(0, \Sigma)$

3

We have $W \sim W_1(n, \Sigma)$. In this instance, $\Sigma = \sigma^2$.

Then Let $X \sim N(\mu, \sigma^2)$. Then,since our degrees of freedom = 1, we have $W = \sum_{i=1}^{n} X^2 = \sum_{i=1}^{n} \frac{\sigma x}{\sigma}\frac{\sigma x}{\sigma} = \sigma^2 \sum_{i=1}^{n} (\frac{x}{\sigma})^2 = \sigma^2 \sum_{i=1}^{n} N(0,1)^2 \sim \sigma^2 \chi_n^2$

We now have our distribution defined as a χ^2 distribution, as desired.

Appendix

```
library(xtable)
options(xtable.floating = FALSE)
options(xtable.timestamp = "")
library(ellipse)
library(sgeostat)
library(mvoutlier)
```

1A

```
data = read.csv("hw2.csv")
n=30
p=2

x1 = c(data$x1)
x2 = c(data$x2)
X = cbind(x1,x2)

xbar1 = mean(x1)
xbar2 = mean(x2)

Xbar = cbind(xbar1,xbar2)
Xbar = c(Xbar)


diff1 = x1 - xbar1
diff2 = x2 - xbar2
diff = cbind(diff1,diff2)

S = t(diff) %*% diff/29
S = cov(X)
```

1B

For B, we will use the following formula to compute the distances : $d^2 = (\vec{X} - \bar{X})'\Sigma^{-1}(\vec{X} - \bar{X})$
Here is a table counting our original (x_1, x_2) observations, along with their respective distance from $\bar{x} = (5.671873, 1.437181)$.

```
md = mahalanobis(X,Xbar,S)

Sinv=solve(S)
#md = diag(t(t(X)-Xbar)%*%Sinv%*%(t(X)-Xbar))


mdtable = rbind(x1,x2,md)
mdtable = as.table(mdtable)
```

1C

```
cutoff = qchisq(.5,p)
count = length(which(md<cutoff))

cat("The Chi-Square value corresponding to a 50% probability contour = ",cutoff,
    "\nThe number of observations that fall within this distance       = ",count,
    "\nThe proportion of observations that fall within the proability  = ",count/n,"")
```

Thus, we can observe that fifty percent of the observed values lie within the fifty percent probability contour. We can visualize this contour probability with an ellipse:

```
 elps <- t(t(ellipse(S, level=0.5, npoints=1000))+Xbar)

lambda = eigen(S)$values
Gamma = eigen(S)$vectors
c=sqrt(cutoff)
factor = c*sqrt(lambda)




plot(X[,1],X[,2],type="n",xlab="x1",ylab="x2",main="50% Probability Contour Ellipse")
index = md < qchisq(0.5,df=p)
points(xbar1,xbar2,pch =8,col="black")
points(X[,1][index],X[,2][index],col="blue",pch=8,bg="blue")
points(X[,1][!index],X[,2][!index],col="red",pch=8,bg="red")

segments(x0 = xbar1,y0 = xbar2,factor[1]*Gamma[1,1]+xbar1,factor[1]*Gamma[2,1]+xbar2, col = "green")
segments(xbar1,xbar2,factor[2]*Gamma[1,2]+xbar1,factor[2]*Gamma[2,2]+xbar2, col = "green")

lines(elps,col="blue")
legend(1.3, 8, legend=c("missed"  , "observed","mean"),
       col=c("red", "blue","black"), pch = 8, cex=0.8)
```

1D

Here is our QQ plot:

```
qqplot(qchisq(ppoints(30),df=p), md, main="", xlab="Theoretical Quantiles", ylab="Sample Quantiles")
qqline(md,distribution=function(x){qchisq(x,df=p)})
```

Here is our χ^2 plot:

```
a =chisq.plot(X,ask=F)
```