Computer class 2 exercises

Richard Wilkinson

March 2 2018

Question 1

Suppose $X_1, \ldots, X_{10} \sim N(\mu, \sigma^2)$, and that we observe data

$$\{2.561, -0.328, 2.607, 3.466, 2.012, 1.293, -2.301, 1.914, 5.779, 1.369\}$$

- Assume that $\sigma = 2$. Use a Monte Carlo test to test the null hypothesis

$$H_0 : \mu = 1$$

against the alternative

$$H_1 : \mu \neq 1.$$

- Now assume that $\mu = 1$. Use a Monte Carlo test to test the null hypothesis

$$H_0 : \sigma^2 = 1$$

against the alternative

$$H_1 : \sigma^2 > 1.$$

Question 2

We will now consider the example from the lecture notes in which we test for randomness in spatial patterns.

Download the data from MOLE, and load the data using the command

```
load(Class1data.Rdata)
```

Locations of 50 points are stored in the vector `spatial`. Check that you can plot the data with the command

```
plot(spatial)
```

Use a Monte Carlo test to test the null hypothesis that the distribution of points is uniform over the unit square.

Hint: The hardest part to this problem is computing the nearest neighbour distances for any particular pattern of points. You should think about how to do this, but in case you are stuck, one approach is given below.

```
nnsum<-function(x){
    n<-length(x[,1])
    mx1<-matrix(x[,1],n,n,byrow=F)
    mx2<-matrix(x[,2],n,n,byrow=F)
    distances<-((mx1-t(mx1))^2+(mx2-t(mx2))^2)^0.5
    distances<-distances+diag(100,n,n)
    # so that we don't pick the zeros on the diagonal
    return(mean(apply(distances,2,min)))
}
```

Question 3

The MASS package contains information on the birth weight of 189 babies. Let's look at the difference between babies born to mothers who smoke and those whose mothers are non-smokers.

```
library(MASS)
attach(birthwt)
bwt.smoke <- bwt[smoke==1]
bwt.nonsmoke <- bwt[smoke==0]
```

Part i)

We'll look below at whether the two groups have the same mean (which is the usual question of interest), but to begin with consider testing the null hypothesis

$$H_0 : \sigma^2_{smoke} = \sigma^2_{nonsmoke}$$

against the alternative that

$$H_1 : \sigma^2_{smoke} < \sigma^2_{nonsmoke}.$$

Note that a useful test statistic to use is

$$T = \frac{s_{smoke}}{s_{nonsmoke}}$$

where s^2 is the usual estimator of variance.

- Use a randomization test to test H_0 vs H_1. What do you conclude?
- What parametric test would you use to test these hypotheses?

Part ii)

Now assume that both distributions are Gaussian.

- Repeat the hypothesis test using a Monte Carlo test.
- Check the Gaussian assumption. Which of these three approaches (Monte Carlo, randomization, or classical tests) do you prefer? Which do you think is most trustworthy?

Part iii)

Use a randomization test to test

$$H_0 : \mu_{bwt.smoke} = 2600$$

against the alternative

$$H_1 : \mu_{bwt.smoke} \neq 2600$$