MATH11400 Statistics 1 2010–11

Homepage http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/

Problem Sheet 3

Remember: when online, you can access the Statistics 1 data sets from an **R** console by typing

```
load(url("http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/stats1.RData"))
```

*1. Let $X_1, \ldots, X_n$ be a random sample from a Geom (θ) distribution, where θ is a single unknown parameter. State the value of $E(X; \theta)$ and hence find the method of moments estimator of the unknown parameter θ.

2. Let $X_1, \ldots, X_n$ be a random sample from a $N(0, \theta^2)$ distribution, where $\theta > 0$ is a single unknown parameter. Find the method of moments estimator of the unknown parameter θ.

3. The data below are thought to come from a Uniform(0,3) distribution.

```
2.08 2.81 0.04 1.54 1.27 0.74
```

Calculate the corresponding expected quantiles of the Uniform(0,3) distribution. Use **R** to plot the sample quantiles against these expected quantiles, and comment on the fit of the distribution to the data.

*4. The data in the `disasters` data set relates to all British coal mining disasters between March 1851 and March 1962 in which 10 or more men were killed. We will study the 120 gaps in days between successive disasters from the start of the series up to January 1889, which we can extract as follows

```
source("http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/disasters.R")
gaps<-disasters$gap[2:121]
```

(if you are not going to be online when using **R**, you have first to copy the file from the downloads section of the website and save it on your computer, then use the `Source R code` item on the `File` menu in **R** to navigate to the saved file and load it in).

(a) Use **R** to plot a histogram of the `gaps` data. Does the histogram indicate that an exponential distribution would or would not be an appropriate model for this set of data? Note that if the occurrence of disasters was completely at random, then the times between disasters would have the 'lack of memory' property, i.e. follow an exponential distribution.

(b) Assuming an Exponential distribution with parameter θ is an appropriate model, show that method of moments estimate of θ for this data set is $\hat{\theta} = 0.008681$.

(c) You are given that the distribution function $F_X(x; \theta)$ has inverse

$$F_X^{-1}(y; \theta) = -\log(1 - y)/\theta$$

Use **R** to produce a quantile plot of the sample quantiles (the ordered observations) against the corresponding approximate expected quantiles for the fitted Exponential distribution and comment on how well the estimated Exponential distribution fits the data. If you prefer to do this exercise without **R** (and this is really **not** recommended!), then I suggest you use a smaller subset of the data, obtained by say

```
gaps<-disasters$gap[2:21]
```

*5. In an experiment to investigate the effect of seeding clouds, the rainfall measurements below were recorded for 25 seeded clouds. The data are contained in the Statistics 1 data set `seeded.rain`.

```
  4.1    7.7   17.5   31.4   32.7   40.6   92.4  115.3  118.3
119.0  129.6  198.6  200.7  242.5  255.0  274.7  302.8  334.1
430.0  489.1  703.4  978.0 1656.0 1697.8 2745.6
```

It is thought that a distribution in the family Gamma(α, λ) may provide a good model for the data. Write down the two equations which determine the method of moments estimates of the two unknown parameters α and λ. Solve these equations to find explicit expressions for $\hat{\alpha}$ and $\hat{\lambda}$, each in terms of the sample moments m_1 and m_2.

6. Continuing from the question above, **R** gives the sample moments for the data as $m_1 = 448.676$ and $m_2 = 623670$ and the resulting method of moments estimates as $\hat{\alpha} = 0.4766319$ and $\hat{\lambda} = 0.001062308$.

   Using the `qgamma` command in **R**, compute the quantiles $F_X^{-1}(k/(n+1); \hat{\alpha}, \hat{\lambda})$ of the fitted Gamma distribution, for $n = 25$ and $k = 1, \ldots, 25$. Use **R** to plot the ordered observations against the approximate expected quantiles for the fitted distribution and comment on how well the fitted model agrees with the data.

7. The following data come from a series of experiments by Henry Cavendish in 1798, designed to measure the density of the earth. They record his measurements of the density of the earth as a multiple of the density of water. The data are contained in the Statistics 1 data set `cavendish`.

```
5.50 5.61 4.88 5.07 5.26 5.55 5.36 5.29 5.58 5.65
5.57 5.53 5.62 5.29 5.44 5.34 5.79 5.10 5.27 5.39
5.42 5.47 5.63 5.34 5.46 5.30 5.75 5.68 5.85
```

(a) Use exploratory methods (histogram, boxplot etc.) to see if there is any immediate reason to believe a $N(\mu, \sigma^2)$ distribution would not provide an appropriate model for these data.

(b) Calculate the method of moments estimates of μ and σ^2. Use the estimates to produce a plot in **R** of the sample quantiles against the corresponding approximate expected quantiles for the fitted Normal distribution. Comment on how well the estimated Normal distribution fits the data.

(c) You can produce an automatic plot of the quantiles of a data set against the corresponding quantiles of a standard $N(0, 1)$ distribution, and, if desired, a fitted line through the first and third quartiles, with the commands

```
> qqnorm(dataset)
> qqline(dataset)
```

where `dataset` is the name of the data file.

Comment on the similarities and differences between this plot and the plot in part (b) above.

[For more information on these two commands, type `help(qqnorm)` and `help(qqline)`]