MATH11400 Statistics 1 2010–11

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

Solution Sheet 3

1. For a Geometric distribution with parameter θ and probability mass function

$$p(x;\theta) = \begin{cases} \theta(1-\theta)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

the population mean is given in terms of the single unknown parameter by $E(X;\theta) = 1/\theta$.

Let m_1 denote the sample mean $\bar{x} = (x_1 + \cdots + x_n)/n$. For simplicity, write $\hat{\theta}$ for the method of moments estimate $\hat{\theta}_{\text{mom}}$. Since there is only one unknown parameter, $\hat{\theta}$ satisfies the single equation

$$\mathrm{E}(X;\hat{\theta}) = m_1 \qquad \text{i.e.} \quad 1/\hat{\theta} = \bar{x} \qquad \text{i.e.} \quad \hat{\theta} = 1/\bar{x} = n/\sum_{i=1}^{n} x_i.$$

2. Since there is only one unknown parameter θ, we need only one equation to define the method of moments estimate, and we use the equation involving the smallest moment of the distribution that explicitly depends on θ.

Usually this would be the first population or distribution moment, $\mathrm{E}(X;\theta)$. However, if X has a $N(0,\theta^2)$ distribution, then the first moment of the distribution is $\mathrm{E}(X;\theta) = 0$ and this does not depend on θ. The next smallest moment is the second moment, $\mathrm{E}(X^2;\theta) = \mathrm{Var}(X;\theta) + [\mathrm{E}(X;\theta)]^2 = \theta^2$ and this does depend on θ.

Let m_2 denote the second sample moment, so $m_2 = (x_1^2 + \cdots + x_n^2)/n$. For simplicity, write $\hat{\theta}$ for the method of moments estimate $\hat{\theta}_{\text{mom}}$. Then the equation defining $\hat{\theta}$ here is

$$\mathrm{E}(X^2;\hat{\theta}) = m_2 \qquad \text{i.e.} \quad \hat{\theta}^2 = \sum_{i=1}^{n} x_i^2/n \qquad \text{i.e.} \quad \hat{\theta} = \sqrt{(}\sum_{i=1}^{n} x_i^2/n).$$

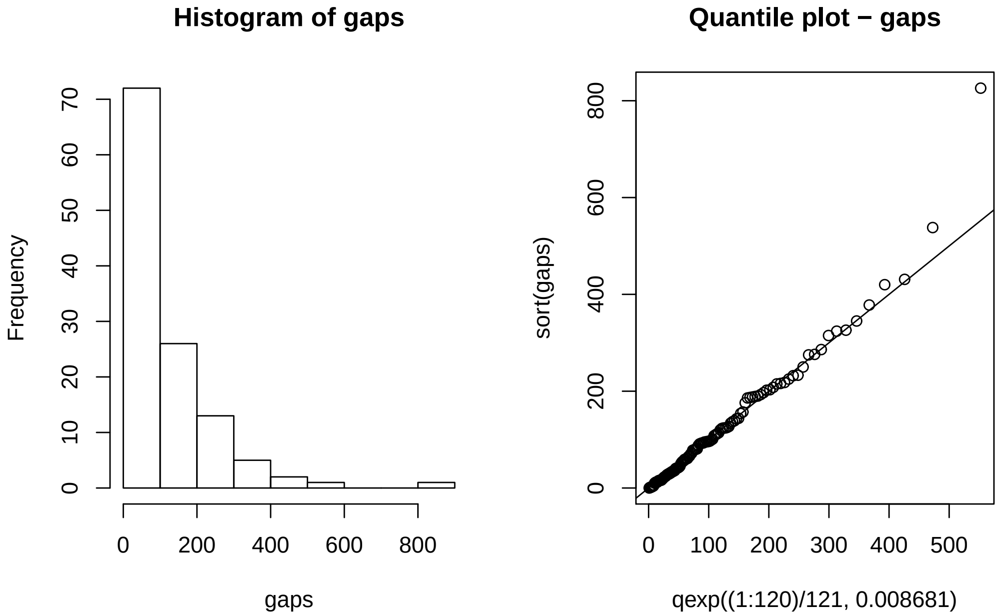
3. (a) The default histogram produced by **R** is shown below. The shape looks like a typical Exponential pdf, so there is no reason to believe an Exponential distribution with parameter θ would not be an appropriate model.

(b) Since there is only one unknown parameter θ, we need only one equation to define the method of moments estimator, and we use the equation involving the smallest moment of the distribution that explicitly depends on θ. For an Exp(θ) distribution, the population mean is $1/\theta$. Let m_1 denote the sample mean $\bar{x} = (x_1 + \cdots + x_n)/n$ and write $\hat{\theta}$ for the method of moments estimator $\hat{\theta}_{\text{mom}}$, then $\hat{\theta}$ satifies the equation

$$\mathrm{E}(X;\hat{\theta}) = m_1 \qquad \text{i.e.} \quad 1/\hat{\theta} = \bar{x} \qquad \text{i.e.} \quad \hat{\theta} = 1/\bar{x}$$

For the given data, $m_1 = 115.2$, giving $\hat{\theta} = 1/m_1 = 0.008680556$.

(c) The steps are: first re-order the observations to obtain the sample quantiles or order statistics $x_{(1)}, \ldots, x_{(n)}$. Then, since there are $n = 120$ observations, calculate $F_X^{-1}(k/121;\hat{\theta}) = -\log(1 - k/121)/\hat{\theta}$ for $k = 1, \ldots, 120$. Finally plot each value of $x_{(k)}$ against the corresponding value of $F^{-1}(k/20;\hat{\theta})$.



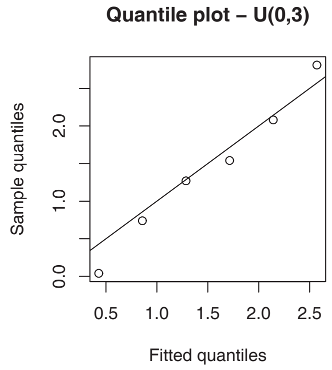
There are many ways of writing a sequence of commands that produce the plot in **R**. The basic steps involved are the same: read in the data; compute the method of moments estimate (here it is given in the question); if necessary, sort the data into ascending order; compute the quantiles of fitted distribution; plot the ordered observations against the fitted quantiles and compare to the line $y = x$. The following basic set of instructions in **R** should make clear exactly what is involved, but note that **R** provides the opportunity for many elegant short cuts once you become familiar with its commands.

```
> m1 <- mean(gaps)
> theta <- 1/m1
> gaps.ord <- sort(gaps)
> quant <- (1:120)/121
> gaps.fit <- qexp(quant,theta) # or: -log(1-quant)/theta
> plot(gaps.fit, gaps.ord,...)  # [add labels and titles here!]
> abline(0,1)
```

The required plot is shown above, on the right. It shows a reasonable fit to the line $y = x$, especially at the lower end, and the only systematic deviation from the line is at the upper end, where the sample quantiles seem somewhat ‘too high’. But there is no very strong reason to believe the observations do not come from an Exponential distribution.

4. For a Unif$(0, 3)$ distribution, $F(x) = x/3$ for the range of interest $0 < x < 3$. Thus $F^{-1}(y) = 3y$, and for the $n = 6$ observations the corresponding five quantiles of the Unif$(0, 3)$ distribution are given by $F^{-1}(k/(6 + 1)) = 3k/7$, for $k = 1, \ldots, 6$. Finally, we plot the ordered observations (the sample quantiles) $x_{(k)}$ against $3k/7$ for $k = 1, \ldots, 6$.

   The resulting plot is shown below, with the straight line $y = x$ added to enable easier assessment of fit. If the observations really came from the given (or fitted) distribution then the points should lie along the this straight line, since we should have $x_{(k)} \simeq F^{-1}(k/7)$. Here the plot shows that there seems to be no systematic deviation from this line. But it is very hard to make any definite judgement with so few data points.



5. (a) Since there are two unknown parameters α and λ, we need two equations to define the method of moments estimators, and we use the equations involving the two smallest moments of the distribution that depend on α and λ. Here $\mathrm{E}(X;\alpha,\lambda) = \alpha/\lambda$ and $\mathrm{E}(X^2;\alpha,\lambda) = \mathrm{Var}(X;\alpha,\lambda) + [\mathrm{E}(X;\alpha,\lambda)]^2 = \alpha/\lambda^2 + \alpha^2/\lambda^2 = \alpha(\alpha+1)/\lambda^2$.

   Let m_1 denote the sample mean $\bar{x} = (x_1 + \cdots + x_n)/n$ and let m_2 denote the second sample moment, so $m_2 = (x_1^2 + \cdots + x_n^2)/n$. For simplicity, write $\hat{\alpha}$ and $\hat{\lambda}$ for the estimators $\hat{\alpha}_{\mathrm{mom}}$ and $\hat{\lambda}_{\mathrm{mom}}$ Then the method of moments estimators satisfy

$$\begin{array}{lrcllrcl} & \mathrm{E}(X;\hat{\alpha},\hat{\lambda}) & = & m_1 & \text{and} & \mathrm{E}(X^2;\hat{\alpha},\hat{\lambda}) & = & m_2 \\ \text{i.e.} & \hat{\alpha}/\hat{\lambda} & = & m_1 & \text{and} & \hat{\alpha}(\hat{\alpha}+1)/\hat{\lambda}^2 & = & m_2 \\ \text{i.e.} & \hat{\alpha}/\hat{\lambda} & = & m_1 & \text{and} & \hat{\alpha}/\hat{\lambda}^2 & = & m_2 - m_1^2 \end{array}$$

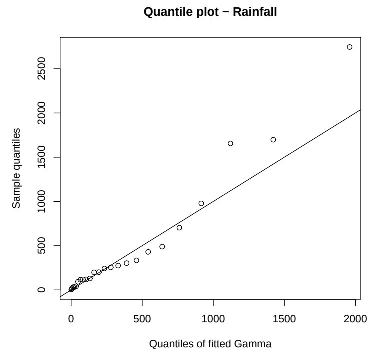
   Whence

$$\hat{\alpha} = \frac{m_1^2}{(m_2 - m_1^2)} \quad \text{and} \quad \hat{\lambda} = \frac{m_1}{(m_2 - m_1^2)}.$$

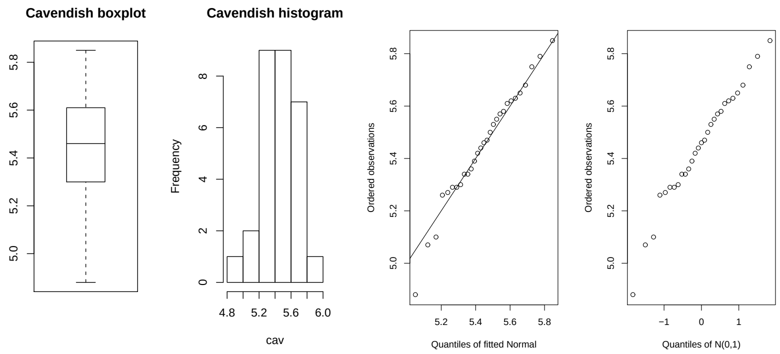
   (b) The method here is again similar to that for question 3, except that there is no closed form calculation which we can use to calculate the inverse of the Gamma distribution function and a package with a command such as the `qgamma` command in **R** is essential. The data here are already ordered, so the following explicit steps would calculate the method of moments estimates from m_1 and m_2 (here they are given in the question) and then find and plot the fitted quantiles.

```
> m1 <- mean(seeded.rain)
> m2 <- mean(seeded.rain^2)
> alpha <- m1^2/(m2 - m1^2)
> lambda <- m1/(m2 - m1^2)
> quant <- seq(1:25)/(26)
> seeded.fit  <- qgamma(quant,alpha,lambda)
> plot(seeded.fit, seeded.rain)
> abline(0,1)
```

   A probability plot for the Rainfall data is given below. The fit does not look exceptionally good; the order statistics in the middle/upper third are somewhat smaller than expected and the three largest observations are much larger than expected. However, I'm not sure that this really shows any systematic deviation from the line, and I feel there is no immediate reason to reject the Gamma model.



6. (a) Both the boxplot and histogram below show no evidence that a Normal distribution would not provide an adequate model for the data. The boxplot indicates that the data are reasonably symmetric about the mean and that there are no outliers, while the histogram looks nicely symmetric and unimodal. A plot (not shown) of the observations in the order given in the question indicates no sign that consecutive observations are not independent.



(b) For the Cavendish data, **R** gives $m_1 = 5.448$ and $m_2 = 29.727$. From your notes, the method of moments estimates are $\hat{\mu} = m_1 = 5.448$ and $\widehat{\sigma^2} = m_2 - (m_1)^2 = 0.047134$, giving $\hat{\sigma} = 0.217$. As in question 1, for a $N(\mu, \sigma^2)$ distribution, $F(x; \mu, \sigma^2) = \Phi((x - \mu)/\sigma)$, where Φ is the distribution function of the $N(0, 1)$ distribution. Thus $F^{-1}(y; \mu, \sigma^2) = \mu + \sigma\Phi^{-1}(y)$, and for the $n = 29$ observations the corresponding quantiles of the fitted distribution are given by $F^{-1}(k/30; \hat{\mu}, \widehat{\sigma^2}) = \hat{\mu} + \hat{\sigma}\Phi^{-1}(k/30)$, for $k = 1, \ldots, 29$.

The method is now the same as for question 3, but using tables of the $N(0, 1)$ distribution or the `qnorm` command in **R**. The probability plot (3rd figure above) shows that the points lie fairly close to the line $y = x$ and shows no systematic deviation from that line, indicating a good fit to the estimated Normal distribution.

(c) From the working in (b) above, we see that the quantiles of the fitted distribution are just $\hat{\mu} +$ ($\hat{\sigma}\times$ the quantiles of the $N(0, 1)$ distribution). Thus we can try plotting the ordered observations against the quantiles of the $N(0, 1)$ distribution, without estimating either μ or σ^2 (see 4th figure above). If the data lie on a straight line in one plot it will lie on a straight line in the other plot, so we can evaluate the fit to a Normal distribution from either plot. Moreover, if we fit a straight line through the points in the $N(0, 1)$ probability plot, we can read off the slope as an estimate of σ and the value at zero as an estimate of μ.