MAST90105 Lab and Workshop 9 Solutions

The Lab and Workshop this week covers problems arising Module 6, Sections 3 to the end. The problems will not be assigned to groups this week.

1 Lab

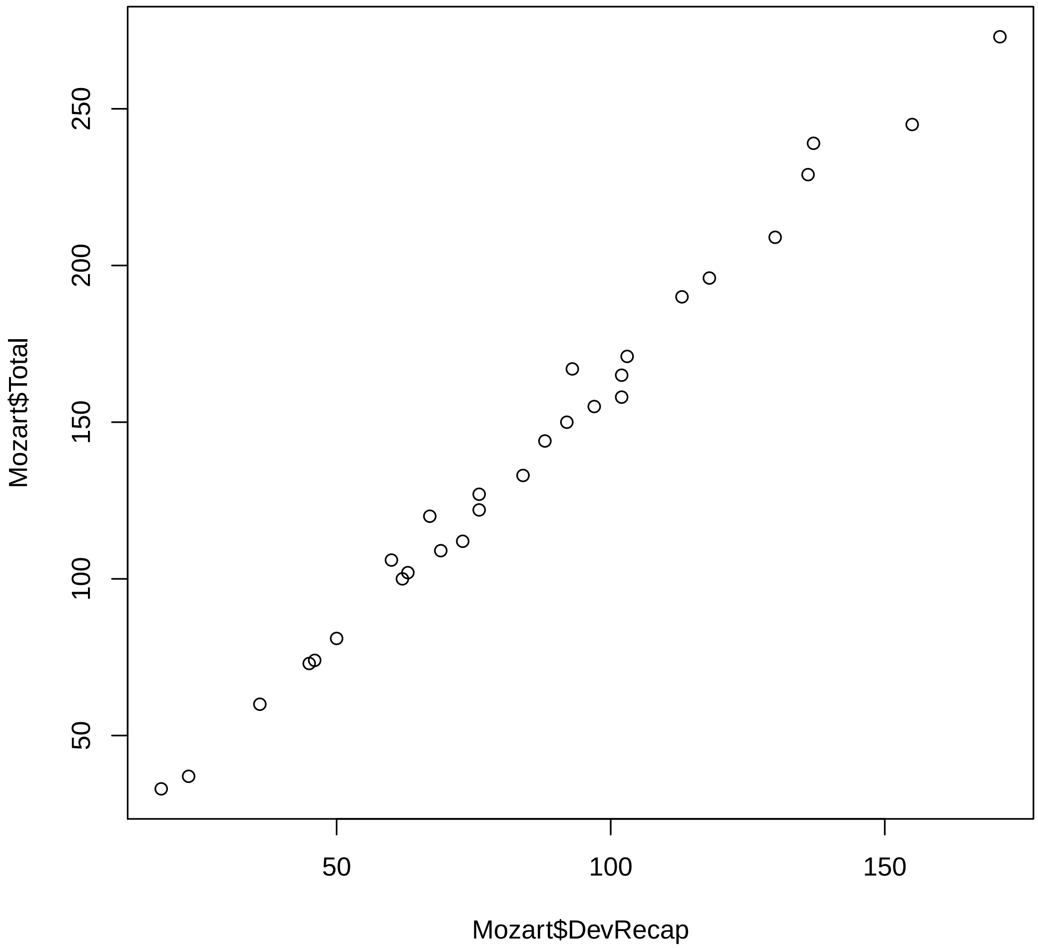
1. (Elaborating Textbook 6.5-10) The "golden" ratio is $\phi = (1+\sqrt{5})/2$. If a, b are numbers, show that they have the "golden" ratio if $\frac{a}{b} = \frac{a+b}{a}$.

   John Putz, a mathematician who was interested in music, analyzed 29 Mozart piano sonata movements which could easily be divided into 2 distinct sections, the Exposition (in which the first and second subjects, the melodies that underly the movement, are revealed) and the Development/Recapitulation (in which the first and second subjects are developed and then restated). Mozart showed interest in mathematics and Putz wondered whether the numbers of bars in the Exposition, b and Development/Recapitulation, a followed the golden ratio (the Recapitulation is often of similar length to the Exposition in Sonata movements from the "classical" period, so that the Development and Recapitulation are always longer than the Exposition).

   The data on the Mozart piano sonata movements is in the lab folder as "Mozart.xls". Import this data into R using the Import Dataset Option.

   a. Make a scatter plot of the points $a+b$ against the points b. Is this plot linear? *The plot looks linear.*

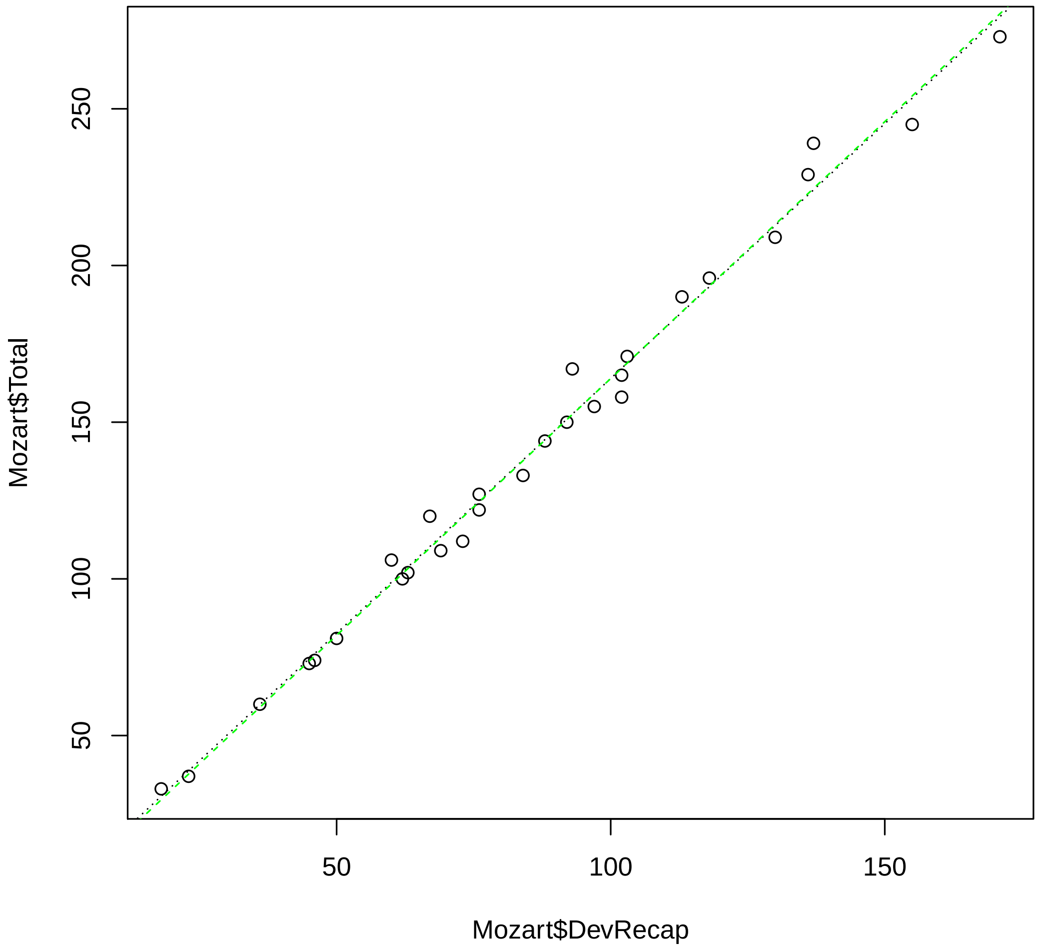
   ```
   # Read Data, Calc. b + a, Plot a + b vs. a
   library(readxl)
   Mozart <- read_excel("L:/MAST90105MethodsofMathematicalStatistics/Mozart.xls")
   Mozart$Total <- Mozart$Expos + Mozart$DevRecap
   plot(Mozart$DevRecap, Mozart$Total)
   #
   ```



b. Find the equation of the least squares regression line with and without intercept. Superimpose them on the scatter plot.

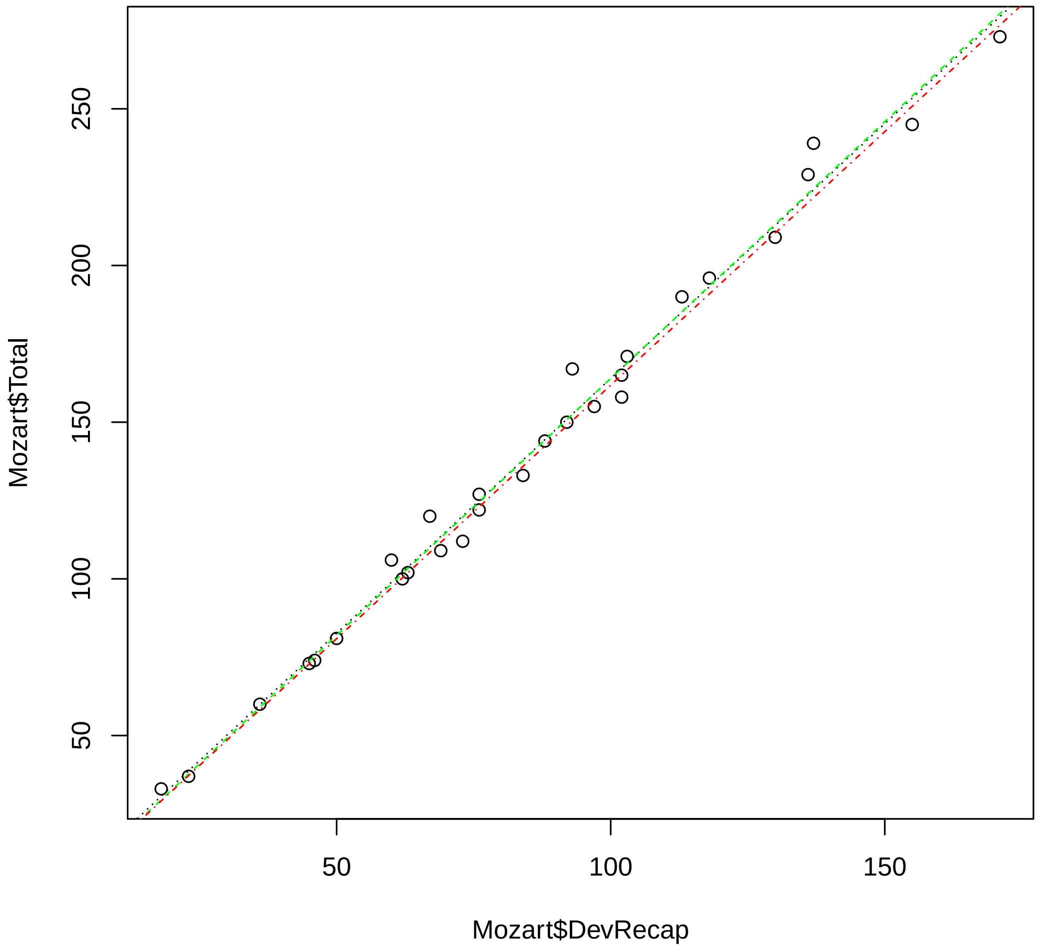
```
# Regr a + b on a with intercept
RegrM <- lm(Mozart$Total ~ Mozart$DevRecap)
summary(RegrM)
# Regr a + b on a without intercept
RegrM0 <- lm(Mozart$Total ~ 0 + Mozart$DevRecap)
summary(RegrM0)
# Add regression line of a + b vs. a, with and
# without intercept
abline(RegrM, lty = "dotted")
abline(RegrM0, lty = "dashed", col = "green")
#
```

```
##
## Call:
## lm(formula = Mozart$Total ~ Mozart$DevRecap)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.210 -3.738 -1.659  2.373 14.880
##
## Coefficients:
##                 Estimate Std. Error t value
## (Intercept)      1.35963    2.88272   0.472
## Mozart$DevRecap  1.62598    0.03085  52.704
##                 Pr(>|t|)
## (Intercept)        0.641
## Mozart$DevRecap   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.19 on 27 degrees of freedom
## Multiple R-squared:  0.9904,Adjusted R-squared:   0.99
## F-statistic:  2778 on 1 and 27 DF,  p-value: < 2.2e-16
##
## Call:
## lm(formula = Mozart$Total ~ 0 + Mozart$DevRecap)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.2115 -4.0148 -0.8182  2.5593 14.5425
##
## Coefficients:
##                 Estimate Std. Error t value
## Mozart$DevRecap  1.63933    0.01213   135.1
##                 Pr(>|t|)
## Mozart$DevRecap   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.104 on 28 degrees of freedom
## Multiple R-squared:  0.9985,Adjusted R-squared:  0.9984
## F-statistic: 1.827e+04 on 1 and 28 DF,  p-value: < 2.2e-16
```



c. On the scatter plot, superimpose the line $y = \phi x$. Compare this line with the least squares regression line.

```
#
Mozart$Golden <- (1 + sqrt(5))/2
abline(a = 0, b = Mozart$Golden[1], lty = "dotdash",
    col = "red")
#
```



Both Least Squares regression lines and the line through the origin with slope the Golden Ratio are very close to the data points and each other.

d. Find the sample mean of the points $(a + b)/b$. Is the mean close to ϕ?

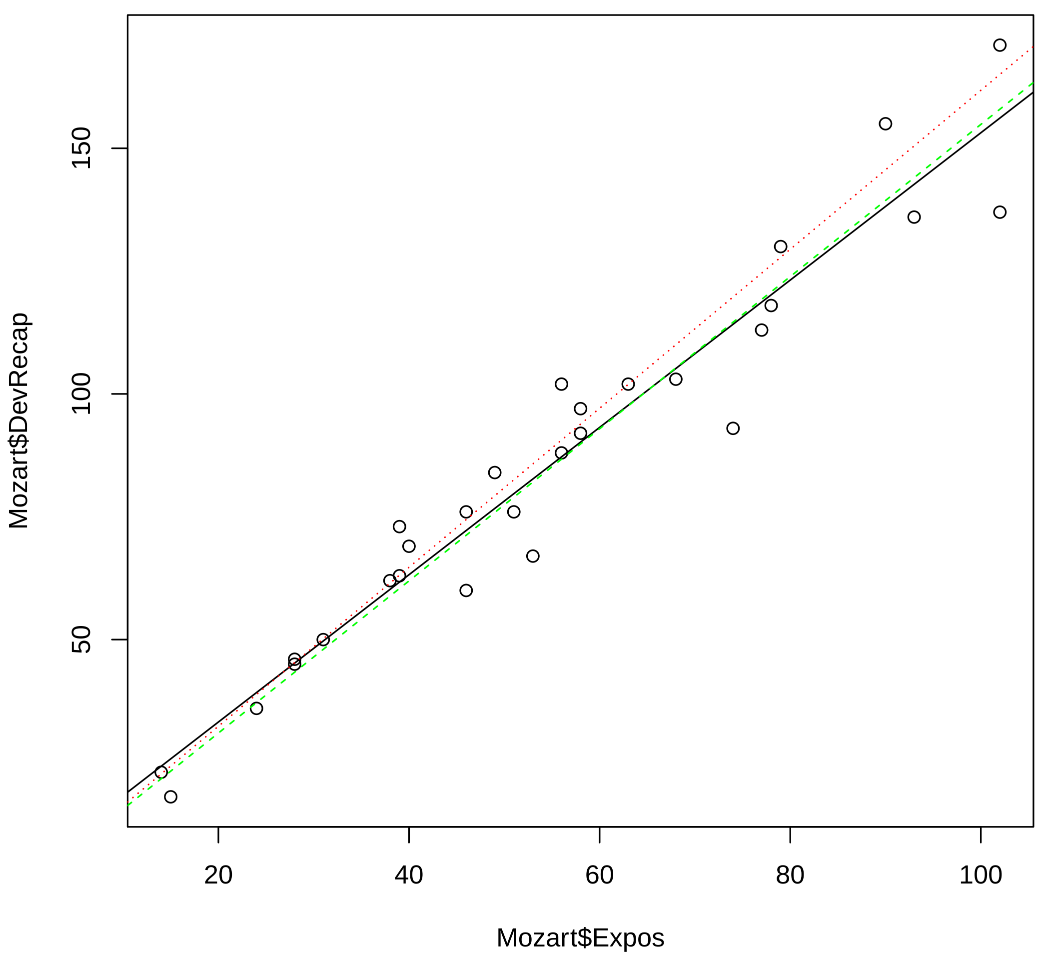
```
# Calculate ratio of a + b to a and find mean
Mozart$Ratio <- Mozart$Total/Mozart$DevRecap
mean(Mozart$Ratio)

## [1] 1.647096

Mozart$Golden[1]

## [1] 1.618034
```

The mean value is close to the golden raio but discrepant by about 0.03, so it is not exact. Notice that the mean of the ratios> least squares slope without intercept >least squares slope with intercept > Golden ratio.

e. Now consider the same questions using the data on a and b. Compare and contrast your results and explain any differences.

```
# Repeat plots and regressions with a vs. b
plot(Mozart$Expos, Mozart$DevRecap)
RegrME <- lm(Mozart$DevRecap ~ Mozart$Expos)
summary(RegrME)

##
## Call:
## lm(formula = Mozart$DevRecap ~ Mozart$Expos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.1738  -3.6930   0.7879   5.7979  16.8395
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.23546    4.43598   0.729    0.472
## Mozart$Expos  1.49917    0.07389  20.290   <2e-16
##
## (Intercept)
## Mozart$Expos ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.579 on 27 degrees of freedom
## Multiple R-squared:  0.9385,Adjusted R-squared:  0.9362
## F-statistic: 411.7 on 1 and 27 DF,  p-value: < 2.2e-16

RegrME0 <- lm(Mozart$DevRecap ~ 0 + Mozart$Expos)
summary(RegrME0)

##
## Call:
## lm(formula = Mozart$DevRecap ~ 0 + Mozart$Expos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.591  -2.975   1.995   7.059  15.632
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## Mozart$Expos  1.54853    0.02938   52.71   <2e-16
##
```

```
## Mozart$Expos ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.499 on 28 degrees of freedom
## Multiple R-squared:   0.99,Adjusted R-squared:  0.9897
## F-statistic:  2778 on 1 and 28 DF,  p-value: < 2.2e-16

abline(RegrME)
abline(RegrME0, lty = "dashed", col = "green")
abline(a = 0, b = (sqrt(5) + 1)/2, lty = "dotted",
    col = "red")
```

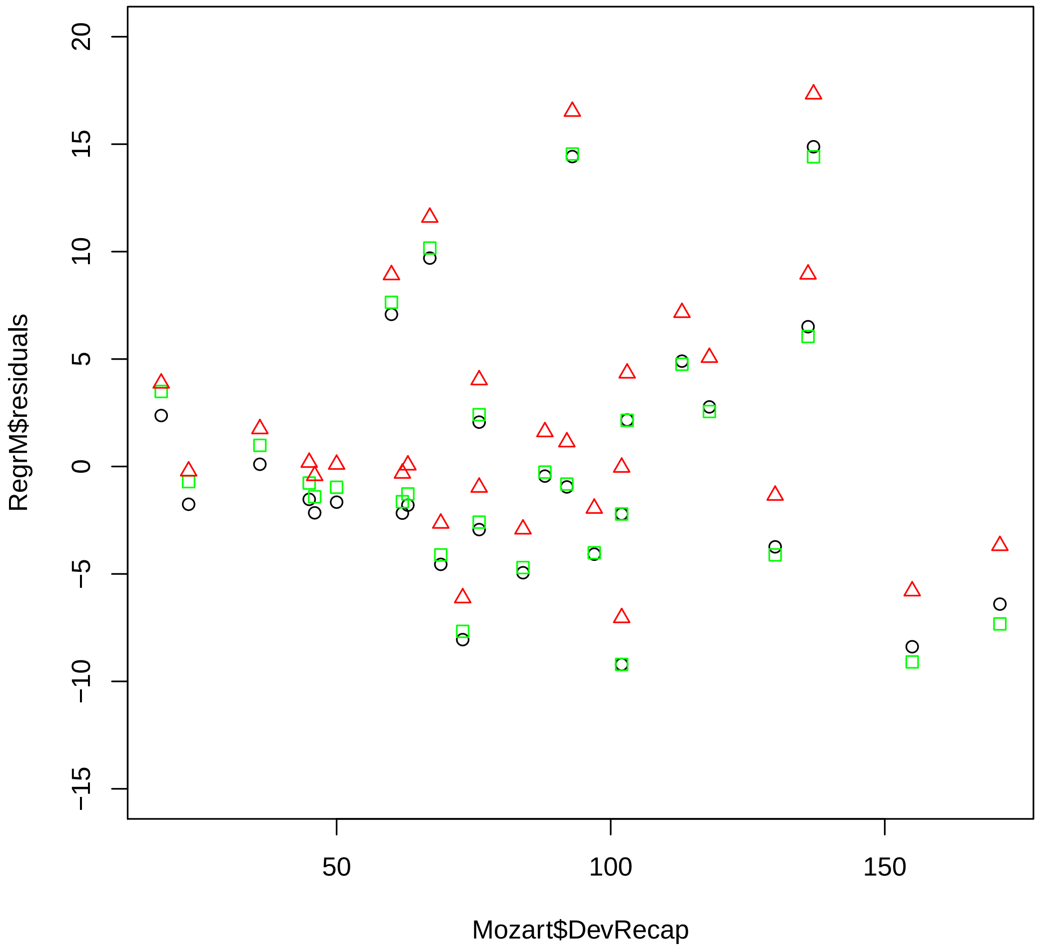


```
# Repeat Ratio Calculations with a/b
Mozart$RatioE <- Mozart$DevRecap/Mozart$Expos
mean(Mozart$RatioE)

## [1] 1.563682

Mozart$Golden[1]

## [1] 1.618034
```

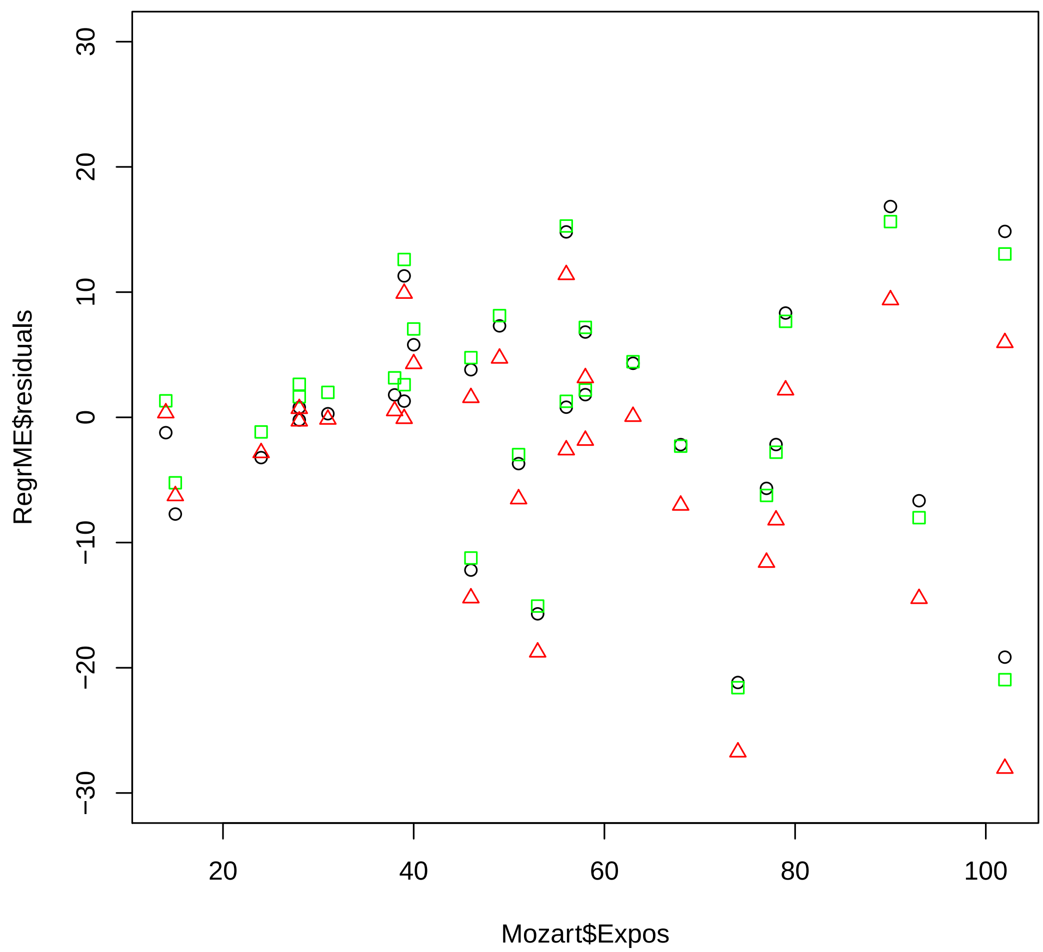
There is much more scatter around both the lines of best fit for the number of bars in the Development/Recapitulation to the number of bars in the Exposition, but these lines are still close to each other. There is a significant difference between the Golden Ratio line and the two lines of best fit.There are a number of sonata movements where the number of bars in the Development/Recapitulation is much lower than the Golden Ratio times the number of bars in the Exposition. This is also observed in the significantly lower value of the mean of the ratios of the number of bars in the Development/Recapitulation to the number of bars in the Exposition. Another contrast is that the order among the least squares slopes and the mean ratio is the same but they are all less than the Golden Ratio.

f. Consider the residuals from the linear models versus the response values as well as the differences between the values from $y = \phi x$ and the response values. In each linear model case, plot the residuals and the difference values on the same plot. Comment on systematic differences

```
# Residual Plots for Total vs.
# Development/Recapitulation
Mozart$GoldenTotal <- Mozart$Golden * Mozart$DevRecap
Mozart$GoldenDiscrep <- Mozart$Total - Mozart$GoldenTotal
plot(Mozart$DevRecap, RegrM$residuals, ylim = c(-15,
    20))
points(Mozart$DevRecap, RegrM0$residuals, col = "green",
    pch = 0)
points(Mozart$DevRecap, Mozart$GoldenDiscrep, col = "red",
    pch = 2)
```




```
# Residual Plots for Development/Recapitulation vs.
# Exposition
Mozart$GoldenDevRecap <- Mozart$Golden * Mozart$Expos
Mozart$GoldenDiscrepE <- Mozart$DevRecap - Mozart$GoldenDevRecap
plot(Mozart$Expos, RegrME$residuals, ylim = c(-30,
    30))
points(Mozart$Expos, RegrME0$residuals, col = "green",
    pch = 0)
points(Mozart$Expos, Mozart$GoldenDiscrepE, col = "red",
    pch = 2)
```



These residuals show very clearly the way in which the scatter is much greater around the regression lines for the Development/Recapitulation number of bars compared to that for the Total. It is also very noticeable that the discrepancies from the golden-ratio times the x-value are all above the residuals for the Total regressions but below them in most cases for the Development/Recapitulation regressions. These systematic differences arise from the fact that the slopes are greater than the golden ratio for the Total regressions but less than the golden ratio for the Development/Recapitulation regressions.

John Putz, in the article that generated this question, proves a mathematical fact, namely that for any numbers $0 \le b \le a$*, if* $\phi = \frac{1+\sqrt{5}}{2}$ *is the Golden Ratio, then* $|\frac{a}{a+b} - \frac{1}{\phi}| \le |\frac{b}{a} - \frac{1}{\phi}|$ *that illustrates why the discrepancies are greater in the Development/Recapitulation than the Total plots.*

g. Do you think Mozart wrote his music thinking about the number of bars in the Development and Recapitulation being the number in the Exposition times the golden ratio? Why?

John Putz gives many references to connections between Mozart's compositions and mathematical patterns and then says in conclusion: "Still, we must remember that these sonatas are the work of a genius, and one who loved to play with numbers. Mozart may have known of the golden section and used it. That there is considerable deviation from it suggests otherwise, however. Perhaps the golden section does, indeed, represent the most pleasing proportion, and perhaps Mozart, through his consummate sense of form, gravitated to it as the, perfect balance between extremes. It is a romantic thought"

So, although the plots show considerable discrepancies in the Development/Recapitulation number of bars from the Golden Ratio times the Exposition number of bars, the similarity in some cases is also striking. The patterns in the music are unquestionable but it is stretching the data too far to say that it proves that he had the Golden Ratio in mind when composing the Development/Recapitulation after the Exposition.

*Mozart was a composer of genius because his compositions can often seem fresh and perfect compared to those of contemporaries and pupils can appear stale and dated today. There is also evidence that Mozart was very keen on Mathematics, for example Marcus du Sautoy (**?**) says: (WebLink) says the following:*

"Mozart loved numbers. Johann Andreas Schachtner, court trumpeter and friend of the Mozart family, wrote about the young Wolfgang: "When he was doing sums, the table, the chair, the walls and even the floor would be covered with chalked numbers."

As an adult Mozart's obsession with numbers didn't wane. He would scatter numbers throughout his letters to family and friends. His family used a secret code to keep politically sensitive comments from the eyes of the censors. But he also used numbers in more intimate exchanges. His kisses would invariably be issued in units of 1,000, although sometimes he would choose a more interesting selection of numbers to shower his correspondent with.

The curious string of numbers 1095060437082 appears in a letter to his wife Constanze. One decoding that has been offered of this sequence suggests we add 10+9+50+60+43+70+82 to get 324, which is 18 squared, again like the opening of Figaro, expressing the bond of love between Mozart and Constanze. He signed himself in another letter as "Friend of the House of Numbers"; while Constanze told a biographer after Mozart's death about "his love of arithmetic and algebra"."

2. Find the maximum likelihood estimator, the Cramér-Rao lower bound and thus the asymptotic variance of the maximum likelihood estimator $\hat{\theta}$ for a random sample $X_1, \ldots, X_n$ taken from the following densities. Determine whether the maximum likelihood estimator is unbiased, and if so, whether the maximum likelihood estimator achieves the lower bound.

   a. $f(x;\theta) = (1/\theta^2)x\exp(-x/\theta)$, $0 < x < \infty$, $0 < \theta < \infty$

      - *Mathematica code and output is in Figure 1.*

```
In[76]:= qn2a = ProbabilityDistribution[1/θ^2 * x * Exp[-x/θ], {x, 0, ∞}, Assumptions → θ > 0]
```

Out[76]= ProbabilityDistribution$\left[\frac{x\, e^{-\frac{x}{\theta}}}{\theta^2}, \{x, 0, \infty\}, \text{Assumptions} \to \theta > 0\right]$

```
h[x_] := Simplify[D[Log[PDF[qn2a, x]], θ], Assumptions → 1 > x > 0]

In[106]:= fisherInformation = Expectation[h[x]^2, x ≈ qn2a]
```

Out[106]= $\frac{2}{\theta^2}$

```
In[78]:= mle[n_] := Solve[Sum[h[x[i]], {i, 1, n}] == 0, θ]

In[82]:= estimates = {mle[2], mle[3], mle[4]}
```

Out[82]= $\left\{\left\{\left\{\theta \to \frac{1}{4}(x[1] + x[2])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{6}(x[1] + x[2] + x[3])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{8}(x[1] + x[2] + x[3] + x[4])\right\}\right\}\right\}$

```
In[96]:= mu = Expectation[θ /. estimates, {x[1] ≈ qn2a, x[2] ≈ qn2a, x[3] ≈ qn2a, x[4] ≈ qn2a}]
```

Out[96]= $\{\{\theta\}, \{\theta\}, \{\theta\}\}$

```
In[108]:= LowerBound = 1/({2, 3, 4} * fisherInformation)
```

Out[108]= $\left\{\frac{\theta^2}{4}, \frac{\theta^2}{6}, \frac{\theta^2}{8}\right\}$

```
In[110]:= varianceMLE =
  (Expectation[(θ /. estimates)^2, {x[1] ≈ qn2a, x[2] ≈ qn2a, x[3] ≈ qn2a, x[4] ≈ qn2a}] -
    mu^2)
```

Out[110]= $\left\{\left\{\frac{\theta^2}{4}\right\}, \left\{\frac{\theta^2}{6}\right\}, \left\{\frac{\theta^2}{8}\right\}\right\}$

Figure 1: Question 2a Mathematica output

- *The MLE is* $\bar{X}/2$.
- *The Fisher information is* $2/\theta^2$.
- *The Cramer-Rao lower bound is* $\theta^2/(2n)$.
- *The MLE is unbiased and achieves the lower bound.*
- *The theoretical calculations are facilitated by recognising the density as Gamma with scale* θ *and index 2.*

b. $f(x;\theta) = (1/(2\theta^3))x^2 \exp(-x/\theta)$, $0 < x < \infty$, $0 < \theta < \infty$

```
In[44]:= qn2b = ProbabilityDistribution[1 / (2 * θ ^ 3) * x ^ 2 * Exp[-x / θ], {x, 0, ∞},
           Assumptions → θ > 0]
```

Out[44]= ProbabilityDistribution$\left[\frac{x^2 e^{-\frac{x}{\theta}}}{2\theta^3}, \{x, 0, \infty\}, \text{Assumptions} \to \theta > 0\right]$

```
In[47]:= h[x_] := Simplify[D[Log[PDF[qn2b, x]], θ], Assumptions → 1 > x > 0]

In[48]:= fisherInformation = Expectation[h[x] ^ 2, x ≈ qn2b]
```

Out[48]= $\frac{3}{\theta^2}$

```
In[49]:= mle[n_] := Solve[Sum[h[x[i]], {i, 1, n}] == 0, θ]

In[50]:= estimates = {mle[2], mle[3], mle[4]}
```

Out[50]= $\left\{\left\{\left\{\theta \to \frac{1}{6}(x[1] + x[2])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{9}(x[1] + x[2] + x[3])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{12}(x[1] + x[2] + x[3] + x[4])\right\}\right\}\right\}$

```
In[52]:= mu = Expectation[θ /. estimates, {x[1] ≈ qn2b, x[2] ≈ qn2b, x[3] ≈ qn2b, x[4] ≈ qn2b}]

Out[52]= {{θ}, {θ}, {θ}}

In[53]:= LowerBound = 1 / ({2, 3, 4} * fisherInformation)
```

Out[53]= $\left\{\frac{\theta^2}{6}, \frac{\theta^2}{9}, \frac{\theta^2}{12}\right\}$

```
In[55]:= varianceMLE =
          (Expectation[(θ /. estimates) ^ 2 , {x[1] ≈ qn2b, x[2] ≈ qn2b, x[3] ≈ qn2b, x[4] ≈ qn2b}] -
            mu ^ 2)
```

Out[55]= $\left\{\left\{\frac{\theta^2}{6}\right\}, \left\{\frac{\theta^2}{9}\right\}, \left\{\frac{\theta^2}{12}\right\}\right\}$

Figure 2: Question 2b Mathematica output

- *Mathematica code and output is in Figure 2.*
- *The MLE is* $\bar{X}/3$.
- *The Fisher information is* $3/\theta^2$.
- *The Cramer-Rao lower bound is* $\theta^2/(3n)$.
- *The MLE is unbiased and achieves the lower bound.*
- *The theoretical calculations are facilitated by recognising the density as Gamma with scale* θ *and index 3.*

c. $f(x;\theta) = (1/\theta)x^{(1-\theta)/\theta}$, $0 < x < 1, 0 < \theta < \infty$

```
In[55]:= varianceMLE =
          (Expectation[(θ /. estimates)^2, {x[1] ≈ qn2b, x[2] ≈ qn2b, x[3] ≈ qn2b, x[4] ≈ qn2b}] -
            mu^2)
```

Out[55]= $\left\{\left\{\frac{\theta^2}{6}\right\}, \left\{\frac{\theta^2}{9}\right\}, \left\{\frac{\theta^2}{12}\right\}\right\}$

```
In[56]:= qn2c = ProbabilityDistribution[θ^-1 x^((1 - θ) / θ), {x, 0, 1}, Assumptions → θ > 0]
```

Out[56]= ProbabilityDistribution$\left[\frac{x^{\frac{1-\theta}{\theta}}}{\theta}, \{x, 0, 1\}, \text{Assumptions} \to \theta > 0\right]$

```
In[59]:= h[x_] := Simplify[D[Log[PDF[qn2c, x]], θ], Assumptions → 1 > x > 0 ]

In[60]:= fisherInformation = Expectation[h[x]^2, x ≈ qn2c]
```

Out[60]= $\frac{1}{\theta^2}$

```
In[61]:= mle[n_] := Solve[Sum[h[x[i]], {i, 1, n}] == 0, θ]

In[62]:= estimates = {mle[2], mle[3], mle[4]}
```

Out[62]= $\left\{\left\{\left\{\theta \to \frac{1}{2}(-\text{Log}[x[1]] - \text{Log}[x[2]])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{3}(-\text{Log}[x[1]] - \text{Log}[x[2]] - \text{Log}[x[3]])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{4}(-\text{Log}[x[1]] - \text{Log}[x[2]] - \text{Log}[x[3]] - \text{Log}[x[4]])\right\}\right\}\right\}$

```
In[63]:= mu = Expectation[θ /. estimates, {x[1] ≈ qn2c, x[2] ≈ qn2c, x[3] ≈ qn2c, x[4] ≈ qn2c}]
```

Out[63]= $\{\{\theta\}, \{\theta\}, \{\theta\}\}$

```
In[64]:= LowerBound = 1 / ({2, 3, 4} * fisherInformation)
```

Out[64]= $\left\{\frac{\theta^2}{2}, \frac{\theta^2}{3}, \frac{\theta^2}{4}\right\}$

```
In[66]:= varianceMLE =
          (Expectation[(θ /. estimates)^2, {x[1] ≈ qn2c, x[2] ≈ qn2c, x[3] ≈ qn2c, x[4] ≈ qn2c}] -
            mu^2)
```

Out[66]= $\left\{\left\{\frac{\theta^2}{2}\right\}, \left\{\frac{\theta^2}{3}\right\}, \left\{\frac{\theta^2}{4}\right\}\right\}$

Figure 3: Question 2c Mathematica output

- *Mathematica code and output is in Figure 3.*
- *The MLE is* $\overline{-\ln X}$.
- *The Fisher information is* $1/\theta^2$.
- *The Cramer-Rao lower bound is* θ^2/n.
- *The MLE is unbiased and achieves the lower bound.*
- *The theoretical calculations involve only rules of expectation and polynomial integration.*

In each case, attempt the questions with and without the aid of Mathematica.

3. (Textbook 6.5-2) In some situations where the regression models is useful, it is known that the mean of Y when $X = 0$ is 0 i.e., $Y_i = \beta x_i + \epsilon_i$ where ϵ_i for $i = 1, \cdots, n$ are independent and distributed as $N(0, \sigma^2)$.

   a. Obtain the maximum likelihood estimators, $\hat{\beta}$ and $\hat{\sigma^2}$, of β and σ^2 under this model.

   - $$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \beta x_i)^2.)$$

     so that

     $$\ell(\beta, \sigma^2) = \ln L(\beta) = C - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta x_i)^2$$

     and

     $$s(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2x_i(Y_i - \beta x_i)$$

     and $s(\beta) = 0$ *when* $\beta = \hat{\beta} = \frac{\sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2}$. *The cross-products cancel as in the standard regression model giving the Analysis of Variance identity leading to the log likelihood expressed as:*

     $$\ell(\beta, \sigma^2) = \ln L(\beta) = C - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((Y_i - \hat{\beta}x_i)^2 + (\hat{\beta}x_i - \beta x_i)^2\right)$$

     . Arguing as in lectures, this maximised if $\beta = \hat{\beta}$ *and* $\sigma^2 = \hat{\sigma^2} = \frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}x_i)^2}{n}$

   b. Find the distributions of $\hat{\beta}$ and $\hat{\sigma^2}$ (You may use, without proof, the fact that $\hat{\beta}$ and $\hat{\sigma^2}$ are independent.)

   - $\hat{\beta}$ *is a linear combination of independent normal random variables so has a normal distribution with mean which is the linear combination of means, and is thus easily seen to be* β, *and variance the sum of the variances divided by the square of* $\sum_{i=1}^{n} x_i^2$, *so is* $\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$. *The Analysis of Variance Identity and the Moment Generating Function arguments relying on the independence of the two estimators give, as in Lectures, the distribution of* $\frac{n\hat{\sigma^2}}{\sigma^2}$ *as* $\chi^2(n-1)$ *(only* β *is estimated).*

   c. By looking up help in R on "formula" find out the R commands to fit this model (recall the we used "lm" in lectures for regression)

      lm (y ~ 0 + x)

4. Let $X_1, \ldots, X_n$ be a random sample from a gamma distribution with $\alpha = 4$ so that

   $$f(x;\theta) = \frac{1}{6\theta^4}x^3 e^{-x/\theta},\ 0 < x < \infty,\ 0 < \theta$$

   Use Mathematica to compute the appropriate derivatives and means.

a. Find the maximum likelihood estimator of θ.

b. Is the MLE an efficient estimator of θ?

c. Give an approximate $100(1-\alpha)\%$ confidence interval for θ.

```
In[1]:= qn4 = ProbabilityDistribution[1 / (6 * θ ^ 4) * x ^ 3 * Exp[-x / θ], {x, 0, ∞}, Assumptions → θ > 0]
```

Out[2]= ProbabilityDistribution$\left[\frac{x^3 e^{-\frac{x}{\theta}}}{6\,\theta^4}, \{x, 0, \infty\}, \text{Assumptions} \to \theta > 0\right]$

```
In[4]:= h[x_] := Simplify[D[Log[PDF[qn4, x]], θ], Assumptions → x > 0]

In[5]:= fisherInformation = Expectation[h[x] ^ 2, x ≈ qn4]
```

Out[5]= $\frac{4}{\theta^2}$

```
In[6]:= mle[n_] := Solve[Sum[h[x[i]], {i, 1, n}] == 0, θ]

In[7]:= estimates = {mle[2], mle[3], mle[4]}
```

Out[7]= $\left\{\left\{\left\{\theta \to \frac{1}{8}(x[1]+x[2])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{12}(x[1]+x[2]+x[3])\right\}\right\}, \left\{\left\{\theta \to \frac{1}{16}(x[1]+x[2]+x[3]+x[4])\right\}\right\}\right\}$

```
In[8]:= mu = Expectation[θ /. estimates, {x[1] ≈ qn4, x[2] ≈ qn4, x[3] ≈ qn4, x[4] ≈ qn4}]
```

Out[8]= $\{\{\theta\}, \{\theta\}, \{\theta\}\}$

```
In[9]:= LowerBound = 1 / ({2, 3, 4} * fisherInformation)
```

Out[9]= $\left\{\frac{\theta^2}{8}, \frac{\theta^2}{12}, \frac{\theta^2}{16}\right\}$

```
In[10]:= varianceMLE =
  (Expectation[(θ /. estimates) ^ 2 , {x[1] ≈ qn4, x[2] ≈ qn4, x[3] ≈ qn4, x[4] ≈ qn4}] - mu ^ 2)
```

Out[10]= $\left\{\left\{\frac{\theta^2}{8}\right\}, \left\{\frac{\theta^2}{12}\right\}, \left\{\frac{\theta^2}{16}\right\}\right\}$

Figure 4: Question 4 Mathematica output

- *Mathematica code and output is in Figure 4.*
- *The MLE is* $\bar{X}/4$.
- *The Fisher information is* $4/\theta^2$.
- *The Cramer-Rao lower bound is* $\theta^2/(4n)$.
- *The MLE is unbiased and achieves the lower bound so is efficient.*
- *The confidence interval question is next week's work and was set by mistake. It will be done next week.*

5. Let X be the number of trials needed to observe the 5th success in a sequence of independent Bernoulli trials. Then X has a negative binomial distribution. Using Mathematica if necessary,

a. What is the mean and variance of X?

   - *Let p be the probability of success on a single trial.*
   - *In the probability half, we showed that $E(X) = 5/p, Var(X) = 5(1-p)/p^2$.*

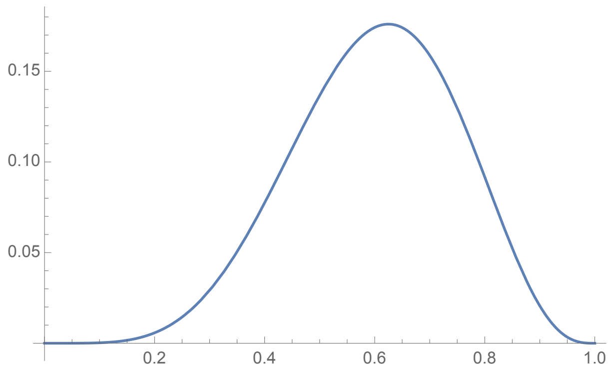
b. Use the Likelihood command with sample size 1 to compute the likelihood of p corresponding to a single observation x on X. Call this L.

```
L = Likelihood[PascalDistribution[5, p], {x}]
```

Out[66]= $\begin{cases} \frac{1}{24} (1-p)^{-5+x} p^5 (-4+x)(-3+x)(-2+x)(-1+x) & x \geq 5 \\ 0 & \text{True} \end{cases}$

In[73]:=
```
Plot[Likelihood[PascalDistribution[5, p], {8}], {p, 0, 1}]
```

Out[73]=



In[74]:=
```
l = LogLikelihood[PascalDistribution[5, p], {x}]
```

Out[74]= $\begin{cases} (-5+x)\,\text{Log}[1-p] + 5\,\text{Log}[p] + \text{Log}\left[\frac{1}{24}(-4+x)(-3+x)(-2+x)(-1+x)\right] & -5+x \geq 0 \\ -\infty & \text{True} \end{cases}$

In[75]:=
```
s = D[l, p]
```

Out[75]= $\begin{cases} 0 & x < 5 \\ \frac{5}{1-p} + \frac{5}{p} - \frac{x}{1-p} & \text{True} \end{cases}$

In[76]:=
```
Solve[s == 0, p]
```

Solve: When parameter values satisfy the condition x < 5, the solution set contains a full–dimensional component; use Reduce for complete solution information.

Out[76]= $\left\{\left\{p \to \text{ConditionalExpression}\left[\frac{5}{x}, x \in \mathbb{R}\right]\right\}\right\}$

In[77]:=
```
i = D[s, p]
```

Out[77]= $\begin{cases} \frac{x}{(-1+p)\,p} - \frac{-5+p\,x}{(-1+p)\,p^2} - \frac{-5+p\,x}{(-1+p)^2\,p} & x \geq 5 \\ 0 & \text{True} \end{cases}$

In[79]:=
```
Expectation[-i, x ≈ PascalDistribution[5, p]]
```

Out[79]= $-\frac{5}{(-1+p)\,p^2}$

Figure 5: Question 5 Mathematica output

- *Mathematica output is in Figure 5.*

c. Plot the likelihood as a function of p if we observe $x = 8$.

- *See Figure 5.*

d. Use the LogLikelihood command with sample size 1 to compute the likelihood of p corresponding to a single observation x on X. Deduce the log-likelihood if $X_1, \ldots, X_n$ are independent observations on X.

- *See Figure 5.*
- *The Log Likelihood based on* $X_1, \ldots, X_n$ *is a function of alone plus* $5\log(p) + \sum_{i=1}^{n} X_i \log(1-p)$

e. Compute the derivative of the log-likelihood. Call this s.

- *See Figure 5.*
- $s = \frac{5}{p} - \frac{\sum_{i=1}^{n}(X_i - 5)}{1-p}$.

f. Fnd the maximum likelihood estimator of p based on a single observation x.

- *See Figure 5.*
- $\hat{p} = \frac{5}{x}$

g. Take a derivative of the score function to determine the observed information $\partial^2 \ln f(x;p)/\partial p^2$. Call this i.

- *See Figure 5*
- *Alternatively, differentiating and taking the negative sign gives* $i = \frac{5}{p^2} + \frac{x-5}{(1-p)^2}$.

h. Compute the expected information.

- *See Figure 5*
- *Fisher Information is* $\frac{5}{(1-p)p^2}$

i. What is the Cramér-Rao lower bound of unbiased estimators of p?

- *Based on a sample size of n, the Cramér-Rao lower bound is* $\frac{1-p)p^2}{5n}$.

6. Suppose that $Y \sim \text{Binomial}(n, p)$ with n known. Find the maximum likelihood estimator of p and determine its asymptotic variance. Determine whether the maximum likelihood estimator is unbiased, and if so, whether the maximum likelihood estimator achieves the lower bound. Check your results in Mathematica.

```
qn6 = BinomialDistribution[n, p]

Out[33]= BinomialDistribution[n, p]

In[39]:= h[x_] := Simplify[D[Log[PDF[qn6, x]], p], Assumptions → n >= x >= 0 ]

In[40]:= fisherInformation = Expectation[h[x] ^ 2, x ≈ qn6]
```

Out[40]= $-\dfrac{n}{(-1+p)\,p}$

```
In[41]:= mle = Solve[h[x] == 0, p]
```

Out[41]= $\left\{\left\{p \to \dfrac{x}{n}\right\}\right\}$

```
In[46]:= mu = Expectation[x / n, x ≈  qn6]

Out[46]= p

In[49]:= Simplify[Expectation[(x / n) ^ 2, x ≈ qn6] - mu ^ 2]
```

Out[49]= $-\dfrac{(-1+p)\,p}{n}$

Figure 6: Question 6 Mathematica output

- *Mathematica commands are6*
- *Hence, the Rao-Cramér lower bound is $p(1-p)/n$.*
- *The Maximum Likelihood Estimate is X/n, as can easily be seen taking logs and differentiating manually.*
- *It is unbiased because $E(\bar{X}) = E(X/n) = p$.*
- *It achieves the lower bound because $Var(X/n) = \frac{p(1-p)}{n}$.*

7. Let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$, where σ^2 is known.

   a. Show that $Y = (X_1 + X_2)/2$ is an unbiased estimator of θ.

   - $E(X_1) = E(X_2) = \theta$ *hence* $E(Y) = E(X_1 + X_2)/2 = \theta$.

   b. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of θ.

   - $f(x;\theta) = 1/\sqrt{2\pi\sigma^2}\exp[-1/(2\sigma^2)(x-\theta)^2]$ *so that*

$$-E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right] = \frac{1}{2\sigma^2}E\left[\frac{\partial^2 (x-\theta)^2}{\partial\theta^2}\right] = \frac{1}{\sigma^2}$$

Hence the Rao-Crámer lower bound is

$$\frac{-1}{nE\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right]} = \frac{\sigma^2}{n}.$$

c. The *efficiency* of an estimator is the ratio of the Cramér-Rao lower bound to the variance of the estimator. What is the efficiency of the estimator in (a)?

- $\mathrm{Var}(Y) = \sigma^2/2$ *so the efficiency is*

$$\frac{\sigma^2/n}{\sigma^2/2} = \frac{2}{n}$$

8. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \theta)$, where μ is known.

a. Show the maximum likelihood estimator of θ is $\hat{\theta} = n^{-1}\sum_{i=1}^n (X_i - \mu)^2$.

- $$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} \exp(-\frac{1}{2\theta}\sum_{i=1}^n (x_i - \mu)^2)$$

so that

$$\ell(\theta) = \ln L(\theta) = C - \frac{n}{2}\ln(\theta) - \frac{1}{2\theta}\sum_{i=1}^n (x_i - \mu)^2$$

and

$$s(\theta) = \frac{\partial \ell(\theta)}{\partial\theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n (x_i - \mu)^2$$

and $s(\theta) = 0$ *when* $\theta = n^{-1}\sum_{i=1}^n (x_i - \mu)^2$.

b. Determine the Cramér-Rao lower bound.

- $f(x;\theta) = 1/\sqrt{2\pi\theta}\exp[-1/(2\theta)(x-\mu)^2]$ *so that*

$$E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right] = E\left[\frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}\right] = \left(\frac{1}{2\theta^2} - \frac{\theta}{\theta^3}\right) = -\frac{1}{2\theta^2}$$

so that the Rao-Crámer lower bound is $2\theta^2/n$.

c. What is the approximate distribution of $\hat{\theta}$ for large n?

- *The approximate distribution of* $\hat{\theta}$ *is* $N(\theta, 2\theta^2/n)$.

d. What is the exact distribution of $n\hat{\theta}/\theta$? Use your knowledge of this distribution to determine if $\hat{\theta}$ attains the Cramér-Rao lower bound?

- *Now* $n\hat{\theta}/\theta \sim \chi^2(n)$ *and* $\mathrm{Var}(\chi^2(n)) = 2n$ *so that* $\mathrm{Var}(n\hat{\theta}/\theta) = 2n$ *or* $n^2\mathrm{Var}(\hat{\theta})/\theta^2 = 2n$ *and rearranging yields* $\mathrm{Var}(\hat{\theta}) = 2\theta^2/n$ *so that* $\hat{\theta}$ *does attain the Rao-Crámer lower bound.*

9. Suppose $X \sim U(0, \theta)$. We take one observation on X.

a. Find the method of moments estimator of θ.

- *$E(X) = \theta/2$ so the method of moments estimator is $\tilde{\theta} = 2X$.*

b. Find the maximum likelihood estimator of θ.

- $$f(x;\theta) = \frac{1}{\theta}, \quad 0 < x < \theta$$

  This is maximised when θ is as small as possible, so the MLE is $\widehat{\theta} = X$.

c. The mean square error of an estimator is $MSE = E(\hat{\theta} - \theta)^2$. Show that

$$MSE = \text{Var}(\hat{\theta}) + \text{bias}_{\hat{\theta}}^2$$

where $\text{bias}_{\hat{\theta}} = E(\hat{\theta}) - \theta$. Which of the estimators in (a) and (b) has the smallest MSE?

- $$\begin{aligned} E(\hat{\theta} - \theta)^2 &= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 + 2E\{[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\} \\ &= \text{Var}(\hat{\theta}) + \text{bias}_{\theta}^2 \end{aligned}$$

  as $E\{[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\} = E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = 0$
- *$E(X) = \theta/2$ and $\text{Var}(X) = \theta^2/12$. Hence*

  $$MSE(\tilde{\theta}) = \frac{4\theta^2}{12} + 0^2 = \frac{\theta^2}{3}$$

  and as $E(\widehat{\theta}) = \theta/2$ so $\text{bias}_{\hat{\theta}} = -\theta/2$,

  $$MSE(\hat{\theta}) = \frac{\theta^2}{12} + \left(\frac{-\theta}{2}\right)^2 = \frac{\theta^2}{3}$$

  Thus both estimators have the same mean square error.

10. Let Y be the sum of the observations from a Poisson distribution with mean θ. Let the prior p.d.f. of θ be gamma with parameters α and β so that

$$f(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}, 0 \le \theta < \infty$$

a. Find the posterior p.d.f. of θ given $Y = y$. (Hint: You should be able to recognise the form of the numerator so only consider the terms that involve θ).

- *Now $Y \sim \text{Poisson}(n\theta)$ so that $f(y|\theta) = \exp(-n\theta)(n\theta)^y/y!$, $y = 0, 1, \ldots$.*

  $$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int_0^\infty f(y|\theta)f(\theta)d\theta}$$

  The numerator is

  $$\frac{\exp(-n\theta)(n\theta)^y}{y!}\frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta} \propto \theta^{y+\alpha-1}\exp[-\theta(n + 1/\beta)]$$

  Thus after normalization this is a gamma with parameters $\alpha + y$ and $1/(n + 1/\beta)$.

b. If the loss function is $[w(y) - \theta]^2$ find the Bayesian point estimate $w(y)$ of θ.

- *Now, we have seen in lectures that the squared error loss function yields the mean of the posterior so* $w(y) = E(\theta|y) = (\alpha + y)/(n + 1/\beta)$.

c. Show that this $w(y)$ is a weighted average of the maximum likelihood estimate y/n and the prior mean $\alpha\beta$, with respective weights $n/(n+1/\beta)$ and $(1/\beta)(n+1/\beta)$.

- $$\frac{\alpha + y}{n + 1/\beta} = \frac{\alpha}{n + 1/\beta} + \frac{y}{n + 1/\beta} = \frac{\alpha\beta}{\beta(n + 1/\beta)} + \frac{y/n}{(n + 1/\beta)/n}$$

  which gives the desired result.