# MAST90105: Lab and Workshop Problems for Week 9

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?
   b. Find the equation of the least squares regression line with and without intercept. Superimpose them on the scatter plot.
   c. On the scatter plot, superimpose the line $y = \phi x$. Compare this line with the least squares regression line.
   d. Find the sample mean of the points $(a + b)/b$. Is the mean close to $\phi$?
   e. Now consider the same questions using the data on $a$ and $b$. Compare and contrast your results and explain any differences.
   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
   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?

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

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

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

In each case, attempt the questions with and without the aid of Mathematica. In using Mathematica, you will need to set up the density function using a function definition ($f[x,\theta] := \cdots$), use the differential operator, D, and expectation using Integrate - consult help for details.

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 $\epsilon_i$ 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 $\beta$ and $\sigma^2$ under this model.

   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)

   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)

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

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

   *Use Mathematica to compute the appropriate derivatives and means.*

   a. Find the maximum likelihood estimator of $\theta$.

   b. Is the MLE an efficient estimator of $\theta$?

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

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 $X$?

   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 $L$.

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

   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 $X$.

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

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

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

h. Compute the expected information.

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

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.

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

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

   b. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of $\theta$.

   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)?

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

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

   b. Determine the Cramér-Rao lower bound.

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

   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?

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

   a. Find the method of moments estimator of $\theta$.

   b. Find the maximum likelihood estimator of $\theta$.

   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?

10. Let $Y$ be the sum of the observations from a Poisson distribution with mean $\theta$. Let the prior p.d.f. of $\theta$ be gamma with parameters $\alpha$ and $\beta$ 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 $\theta$ given $Y = y$. (Hint: You should be able to recognise the form of the numerator so only consider the terms that involve $\theta$).

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

    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 $(1/\beta)(n + 1/\beta)$.