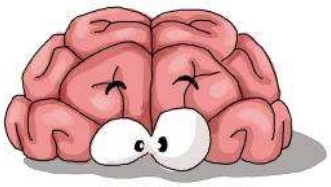
WST211
Practical 6



Date: 15 April 2016
Due: 25 April 2016

Instructions:

- **Answer all questions.**
- **Hand in a typed document with your answers which includes the following:**
- **Table of contents**
- **Answers**
- **SAS program**
- **SAS output**
- **Attach a copy of the SAS program in an appendix.**
- **Attach the relevant outputs of the SAS output.**
- **According to the question, make interpretations about the SAS output.**
- **Round the answers to 3 decimal places.**
- **Hand in a typed document with your answers and include the SAS programs.**

Questions

1. The survival times (in days) of a white rat that was subjected to a certain level of X-ray radiation is a random variable X ~ GAM (5,4). Generate 10000 values from this distribution by using the **RANGAM** function and a seed value of 42. Use the data to estimate the values of the following.

   (a) The expected survival time, E(X).
   (b) Standard deviation of X.
   (c) P ($15 < X \leq 20$).

2. Let X ~ GAM(θ,κ) with θ fixed at 0.5 and κ assuming three different values; 0.5, 1 and 2. Use the PDF function in SAS to calculate the probability density $f_X(x)$ for the three distributions (GAM(0.5,0.5), GAM(0.5,1) and GAM(0.5,2)). Use *x* values from 0 to 3 by 0.1. Plot the three density functions on the same graph. The graph should be similar to Figure 3.3 in Bain and Engelhardt (page 117). Let the scale of both the horizontal and vertical axes be from 0 to 3. What is the effect of the parameter κ on the density function?

   The code for the graph is given below.

```
goptions reset=global;
        symbol1 color=blue interpol=join value=triangle;
        symbol2 color=green  interpol=join value=dot;
        symbol3 color=black  interpol=join value=square;

        proc gplot;
        plot fx1*x fx2*x fx3*x/vaxis=0 to 3 overlay legend;
        run;
```

3. Repeat Question 2 but with $\kappa = 2$ and $\theta = 0.5$, 0.8 and 1.1 respectively. Let the scale of the horizontal axis be from 0 to 5 and for the vertical axes be from 0 to 1. What is the effect of the parameter θ on the density function?

4. The following case study was obtained from the website http://lib.stat.cmu.edu.
Forbes magazine published data on the best small firms in 1993. These were firms with annual sales of more than \$5 million and less than \$350 million. Firms were ranked by five-year average return on investment. The data extracted is the annual salary of the chief executive officer for the first 60 ranked firms. The distribution of the salaries are considered. The data is given in the file salaries.xls on ClickUP.

   Study Example 4.22 from SAS Help given in the document 'Practical 6 Notes' and fit a Gamma, Weibull and Normal distribution to the salaries data. The estimation method is maximum likelihood.

   (a) Plot a histogram of the data and overlay the fitted distributions. The code for the graph is given below.

```
proc univariate;
var salary;
histogram / midpoints=100 to 1000 by 100
     gamma (color=blue l=1)
     weibull (color=red l=20)
     normal (color=blue l=34);
inset n mean(5.3) std='Std Dev'(5.3)
      gamma weibull normal
      /pos = ne  header = 'Summary Statistics';
axis1 label=(a=90 r=0);
run;
```

   (b) Use the Cramer-von Mises to test the hypothesis that the data comes from a Gamma or Normal distribution. Use a significance level of 10% ($\alpha = 0.10$).

   As an example the hypothesis test is done for the Weibull distribution based on the SAS output given below.

   H_0: The data follows a Weibull distribution
   H_1: The data does not follow a Weibull distribution

   Since p-value = 0.037< 0.10 the null hypothesis is rejected. We reject the null hypothesis that the data follows a Weibull distribution at a 10% significant level.

The UNIVARIATE Procedure
Fitted Weibull Distribution for salary

Parameters for Weibull Distribution

Parameter	Symbol	Estimate
Threshold	Theta	0
Scale	Sigma	469.2491
Shape	C	1.854611
Mean		416.7521
Std Dev		233.1886

Goodness-of-Fit Tests for Weibull Distribution

Test		----Statistic-----		------p Value------
Cramer-von Mises	W-Sq	0.13177100	Pr > W-Sq	0.037
Anderson-Darling	A-Sq	0.76504792	Pr > A-Sq	0.044

(c) Which of the three distributions fits the data best? Give the parameter estimates for the distribution that fits the data best.