1. When analysing poetry writing styles, one approach is to count the frequency distribution of the number of words required to reach the next word not already used before in the poem. For instance, in the sentence *There is gold in there but it is too deep* the first four words all have a delay of 0 until a new word is reached, but there is a delay of at the second *there*, which is not new in the sentence. There is a similar delay of 1 at the second *is*. So the sequence of delays for this sentence is 0,0,0,0,1,0,1,0. For the following poem by Robert Frost (1874-1963), *The Road Not Taken*, count the frequencies of waiting times until new words, and plot the histogram. Compute the average waiting time, and use its reciprocal as an estimate of the parameter of an exponential distribution. Construct an exponential quantile-quantile plot using this parameter estimate.

   The Road Not Taken

   Two roads diverged in a yellow wood,
   And sorry I could not travel both
   And be one traveler, long I stood
   And looked down one as far as I could
   To where it bent in the undergrowth;

   Then took the other, as just as fair,
   And having perhaps the better claim,
   Because it was grassy and wanted wear;
   Though as for that the passing there
   Had worn them really about the same,

   And both that morning equally lay
   In leaves no step had trodden black.
   Oh, I kept the first for another day!
   Yet knowing how way leads on to way,
   I doubted if I should ever come back.

   I shall be telling this with a sigh
   Somewhere ages and ages hence:
   Two roads diverged in a wood, and I-
   I took the one less traveled by,
   And that has made all the difference.

2. The following data are errors (measured in degrees towards a 90 degree angle) made by 36 children when trying to draw a 45 degree angle to a

horizontal line.

16	-34	2	18	0	25	-20	3	3	19	55	121
-14	-25	7	114	50	-9	12	10	-10	7	13	7
7	3	8	4	-15	-4	15	17	6	8	32	28

(a) Make a histogram of the data. Would you conjecture the data to be normally distributed?

(b) Make a boxplot of the data. Are there any outliers?

(c) Calculate mean and standard deviation for the sample.

(d) Would there be any indication of a tendency for the children to make a particular type of error?

3. Suppose that $U_1, \ldots, U_n$ is a random sample from the uniform distribution on $(0, 1)$. Let $U_{(k)}$ denote the kth order statistics in the sample, in other words, $U_{(k)}$ is the kth smallest value. Show that

$$EU_{(k)} = \frac{k}{n+1}$$

and

$$VarU_{(k)} = \frac{k}{(n+1)(n+2)} \left(1 - \frac{k}{n+1}\right).$$

Define the median of the random sample, distinguishing the two cases, n even and n odd. Show that the median has expected value $\frac{1}{2}$ and find its variance when n is odd. Explain (without doing it) how you would find the variance in the case n even.

4. Show that if X and Y are independent exponential random variables with mean 1, then $\frac{X}{Y}$ follows an F-distribution. Identify the degrees of freedom.

5. Let $X_1, \ldots, X_n$ be independent and identically $\mathcal{N}(\mu, \sigma^2)$-distributed. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Derive the moment-generating function of S^2 and deduce that $ES^2 = \sigma^2$ and $VarS^2 = 2\sigma^4/(n-1)$. For each case, find a function of $(\bar{X}, S^2, \mu, \sigma^2)$ having the following distributions.

a) standard normal $\mathcal{N}(0, 1)$

b) $\chi^2(n-1)$

c) $\chi^2(n)$

d) $t(n-1)$.