Introduction to Lisp-Stat and its applications.

Exercises 1.

1. RANDU was the name of a random-number generator which was used on IBM mainframe computers some time ago. The claim was that it produces uniform(0,1) random nmkbers. The code below defines a function which implements this generator.

```
(defun randu (n &optional (seed 12345))
     (let ((a (+ (^ 2 16) 3))
           (m (^ 2 31)))
       (flet ((mod (x m) (- x (* m (floor (/ x m))))))
          (do ((i 0 (1+ i))
               (res (list seed)
                    (cons (mod (* a (first res)) m) res)))
              ((>= i n)
               (/ (rest (reverse res)) m))))))
```

   (a) Create a list of random numbers with this generator. Do they look like they come from the correct distribution?

   (b) If the generator works then we hope that there will be no correlation between successive draws. The following code produces 300 generated values and then splits them into three lists with corresponding elements of the three lists being consecutive draws.

```
(def randu-lists
     (apply #'mapcar #'list (split-list (randu 300) 3)))
```

   Examine a `spin-plot` of these lists to see if you can spot the problem with the generator.

2. Use `(load-data "book")` to install the data used in Tierney (1990). The variables `flow`, `moist` and `ratio` arose in an experiment to examine the relationship between the flow rate of sulphate of ammonia crystals in a packing process, the initial moisture content and the shape of the crystals measured as the ratio length/width. Use some of the plotting techniques we have seen to graphically examine these data. Linking of different plots may be helpful.

3. The data wbc-pos and times-pos contain a data set used to study the survival time (in weeks) of leukaemia patients and their white blood cell count at entry. A possible model for this data is that the survival times are exponentially distributed with the log of the mean survival time linear in the log of the blood cell count. That is

$$\mu(x) = \theta_0 \exp(-\theta_1 x)$$

where x is the log of the WBC count. For convenience we shall scale the counts by 10,000 to get transformed-wbc-pos. The log-likelihood for this data is

$$\sum_{i=1}^{n} \theta_1 x_i - n \log(\theta_0) - \frac{1}{\theta_0} \sum_{i=1}^{n} y_i e^{\theta_1 x_i}$$

```
(defun llik (theta)
   (let* ((x transformed-wbc-pos)
          (y times-pos)
          (n (length x))
          (theta0 (select theta 0))
          (theta1 (select theta 1)))
     (- (sum (* theta1 x))
        (* n (log theta0))
        (/ (sum (* y (exp (* theta1 x))))
           theta0))))
```

(a) Use regression-model to get some reasonable starting values for the parameters. Store these in the variable initial-vals

(b) Use (newtonmax #'llik initial-vals) to get the maximum likelihood estimates of the parameters. Confirm that the same results are obtained using nelmeadmax.

4. We could also conduct a Bayesian analysis of this data set examined in the previous question.

   (a) Initially let the prior be uniform on both parameters. Then the log-posterior is equal to the log-likelihood. Use `bayes-model` to conduct the Bayesian analysis of this model and store the result in `leuk-bayes1`.

   (b) `(send leuk-bayes1 :moments)` will return second order approximations to the posterior mean and their standard errors. Are these very different to the first order approximations.

   (c) Plot the two marginal posterior densities which can be extracted using the `:margin1` message.

   (d) Finally suppose we wish to incorporate prior information. θ_0 can be interpreted as the average survival time for a patient with WBC of 10,000. Let us give a normal prior to the log of this parameter with mean log(52) and standard-deviation log(2). θ_1 is the approximate percent difference in mean survival time for patients whose WBC differs by 1%. Let us give this a prior distribution which is normal with mean 5 and standard error 2. Assume that the two parameters are independent apriori. The functions defining the two priors are

```
(defun prior0 (theta0)
   (normal-dens (/ (- (log theta0) (log 52)) (log 2))))
(defun prior1 (theta1)
   (normal-dens (/ (- theta1 5) 2)))
```

   A function to calculate the log posterior would then be

```
(defun lpost (theta)
   (+ (llik theta)
      (log (prior0 (first theta)))
      (log (prior1 (second theta)))))
```

   Examine how this prior information has changed the Bayesian analysis.