Stat-6207 **Homework Set 3** **Fall 2018(Modarres)**

1) To measure the breaking strength of material Swedish physicist Waloddi Weibul proposed the Weibul distribution with density function $f(x) = \lambda x^{\lambda-1}\exp(-x^{\lambda})$ for $x, \lambda > 0$. Suppose you have a random sample $x_1, \ldots, x_n$ from the Weibul distribution and would like to estimate the maximum likelihood of λ.

- a. Find the likelihood function $L(\lambda|x_1, \ldots, x_n)$ and the log-likelihood function.
- b. Find the score functions $S(\lambda) = \frac{\partial \ln(L(\lambda|x_1,\ldots,x_n))}{\partial \lambda}$ and its derivative.
- c. Suppose we observe 5 values $(0.1, 0.25, 0.5, 1, 2)$. Plot the log-likelihood function.
- d. Write an R function that uses Newton's method to find the MLE of λ.

2) Insert keys $(45, 24, 13, 0, 7, 23, 30, 43, 32)$ into a binary search tree where each node is a list of 4 fields: (left pointer,key, rank, right pointer). The rank of a node is one plus the number of nodes in its left subtree.

- a. Show the tree with nodes composed of value and rank field.
- b. Show the inorder traversal of the tree.
- c. Show the sequence of calls to find the median, minimum and maximum.
- d. Let $S[x]$ be the number of nodes in a binary search tree rooted at x. Find a recursive expression for $S[x]$.

3) Suppose the postorder traversal of a binary tree gives $(A, B, G, K, H, F, Y, Z, X, R, M)$ whereas its inorder traversal gives $(A, B, F, G, H, K, M, R, X, Y, Z)$. Draw the binary tree and give its preorder traversal.

4) Obtain the time complexity of the following code segments in terms of N.

- a. $for\ \ (i\ in\ 1:N)\{for\ \ (j\ in\ 1:i)\ print(x)\}$
- b. $f2 = function(N)\{if(N == 0)return(1); x = 0; for\ (i\ in\ 1:N)\ x = x + f2(N-1)\}$
- c. $f3 = function(N)\{if(N == 1)return(0); return(1 + f3(N/2)\}$

5) When a data set is too large to be stored in primary memory, standard algorithms for computing sample quantiles are not directly applicable. You will need to write a function for estimating the sample median of n observations in m memory locations using n random numbers from a standard normal distribution. To describe the method of recursive medians, suppose that m is of the form b^k where b and k are integers. The **remedian algorithm** processes the observations sequentially in groups of size b. A median is computed for each group, yielding b^{k-1} medians in the first stage. This step is repeated recursively until a single estimate is found. A median can be computed using k arrays of size b requiring $m = kb$ storage locations. Let $b = 15$ and $k = 4$ so that $n = 50,625$ and $m = 60$. Replicate your result 100 times. Obtain univariate summary statistics and compare to the true median. See the paper on the Remedian.

6) We are given n objects in a knapsack. Object i has weight w_i and the knapsack has capacity M. If object i is placed into the knapsack, then a profit of p_i is earned. The objective is fill the knapsack to maximize profit. Hence we want to maximize $\sum_{i=1}^{n} p_i x_i$ subject to $\sum_{i=1}^{n} w_i x_i \leq M$ and $x_i = 0, 1$ for $i = 1, \ldots, n$.

- a. Consider the following instance of the knapsack problem: $n = 3$, $M = 20$, $P = (p_1, p_2, p_3) = (25, 24, 15)$, and $W = (w_1, w_2, w_3) = (18, 15, 10)$. By hand calculations find the feasible solutions and obtain the optimal selection when $X = (x_1, x_2, x_3)$ is i) $(1, 0, 1)$, ii) $(1, 0, 0)$, iii) $(0, 0, 1)$, and iv) $(1, 1, 1)$.

- b. We would like to devise a randomized algorithm to find a suboptimal solution. Normally, P and W are given vectors, and we generate X to maximize the objective function. In this case, suppose we generate $n = 20$ observations from $P \sim \chi^2(1)$, and $W \sim U(0, 1)$ and fix them. Write an R function that finds a suboptimal solution by generating 1000 instances of the problem generating vector X with independent components $Bernoulli(0.5)$. Let $M = 2$.