STAT 604 Final Exam Solution **Fall, 1994 Dr. Newton**

1. (10 points) Given the `real*4 function runif(dseed)` that generates a single $U(0,1)$ for input and output double precision seed `dseed`, write `subroutine rlogist(dseed,n,alpha,beta,x)` which generates n observations $X_1, \ldots, X_n$ from the logistic distribution

$$f(x;\mu,\beta) = \frac{1}{\beta}\frac{e^{-(x-\alpha)/\beta}}{[1+e^{-(x-\alpha)/\beta}]^2}, \qquad x \in \mathcal{R}.$$

SOLUTION:

Need to show that quantile function is

$$Q(u) = \alpha - \beta \log\left(\frac{1-u}{u}\right),$$

and thus the subroutine is

```
      subroutine rlogist(dseed,n,alpha,beta,x)

      double precision dseed
      real*4 runif
      dimension x(n)

      do 10 i=1,n
      ui=runif(dseed)
 10   x(i)=alpha-beta*log((1.-u)/u)

      return
      end
```

---

2. (15 points) For the gamma density having parameter $\nu > 1$

$$f_\Gamma(x) = \frac{x^{\nu-1}e^{-x}}{\Gamma(\nu)}, \qquad x > 0,$$

find a double exponential pdf

$$f_{DE}(x;\mu,\sigma) = \frac{1}{2\sigma}e^{-|x-\mu|/\sigma}, \qquad x \in \mathcal{R},\ \mu \in \mathcal{R},\ \sigma > 0$$

having the same mode as f_Γ and a constant c so that $h(x) = cf_{DE}(x) \geq f_\Gamma(x)$ for all x and c is as small as possible. How would you use this to generate gamma data using rejection?

SOLUTION:

Solving $f'_\Gamma(x) = 0$ and checking second derivative shows mode of gamma is $\nu - 1$. Since DE is symmetric about $x = \mu$, we have that the mode of DE is μ and thus we let $\mu = \nu - 1$ in the DE. If we let $g(x) =$

$f_\Gamma(x)/f_{DE}(x)$ and find $c = \max g(x)$, we will have $cf_{DE}(x) \geq f_\Gamma(x)$ for all x. If we let $a = 2\sigma/\Gamma(\nu)$, we have

$$g(x) = \begin{cases} 0, & x \leq 0, \\ g_1(x) = ax^{\nu-1}e^{-x-(\nu-1-x)/\sigma}, & 0 < x \leq \nu - 1, \\ g_2(x) = ax^{\nu-1}e^{-x-(x-\nu+1)/\sigma}, & x \geq \nu - 1. \end{cases}$$

Now g_1 is maximized at $x_1 = \dfrac{\sigma(\nu-1)}{\sigma-1}$ while g_2 at $x_2 = \dfrac{\sigma(\nu-1)}{\sigma+1}$ and we can let $c = \max(g_1(x_1), g_2(x_2))$.

Note that to use rejection, we actually need to consider the DE truncated at zero, that is, use f_{DE} divided by its integral from $-\infty$ to 0.

---

3. (15 points) Let $\tilde{X}_n$ be the median of $n = 2k+1$ distinct observations $X_1, \ldots, X_n$. Show that the jackknifed version of $\tilde{X}_n$ (the average of the n "leave one out" medians) is given by

$$\hat{X} = \frac{k}{n}\tilde{X}_n + \frac{k+1}{2n}\left[X_{(k)} + X_{(k+2)}\right],$$

where the order statisics of the data are denoted by $X_{(1)} < \cdots < X_{(n)}$.

SOLUTION:

If we let $\tilde{X}_{(i)}$ denote the median of the X's with $X_{(i)}$ left out, we have

$$2\tilde{X}_{(i)} = \begin{cases} X_{(k+1)} + X_{(k+2)}, & i = 1, \ldots, k, \\ X_{(k)} + X_{(k+2)}, & i = k+1, \\ X_{(k)} + X_{(k+1)}, & i = k+2, \ldots, n, \end{cases}$$

and thus

$$\hat{X} = \frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{(i)} = \frac{1}{2n}\left[2kX_{(k+1)} + (k+1)X_{(k)} + (k+1)X_{(k+2)}\right],$$

which gives the result.

---

4. (15 points) Find a recursive method for finding the sample variance s_n^2 of a sample $X_1, \ldots, X_n$, that is, a formula which finds s_n^2 as a function only of s_{n-1}^2, $\bar{X}_{n-1}$, and X_n.

SOLUTION:

There are many ways to get the recursion but one simple way is to write

$$\begin{aligned}(n-1)s_n^2 - (n-2)s_{n-1}^2 &= \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 - \sum_{i=1}^{n-1}(X_i - \bar{X}_{n-1})^2 \\ &= \sum_{i=1}^{n}\left[(X_i - \bar{X}_{n-1}) - (\bar{X}_n - \bar{X}_{n-1})\right]^2 - \sum_{i=1}^{n-1}(X_i - \bar{X}_{n-1})^2, \\ &= \sum_{i=1}^{n}(X_i - \bar{X}_{n-1})^2 + n(\bar{X}_n - \bar{X}_{n-1})^2 - 2(\bar{X}_n - \bar{X}_{n-1})\sum_{i=1}^{n}(X_i - \bar{X}_{n-1}) - \sum_{i=1}^{n-1}(X_i - \bar{X}_{n-1})^2 \\ &= (X_n - \bar{X}_{n-1})^2 - n(\bar{X}_n - \bar{X}_{n-1})^2,\end{aligned}$$

since $\sum_{i=1}^{n}(X_i - \bar{X}_{n-1}) = \sum_{i=1}^{n} X_i - n\bar{X}_{n-1} = n\bar{X}_n - n\bar{X}_{n-1} = n(\bar{X}_n - \bar{X}_{n-1})$.

Now $n\bar{X}_n = \sum_{i=1}^{n-1} X_i + X_n = (n-1)\bar{X}_{n-1} + X_n$ and thus $n(\bar{X}_n - \bar{X}_{n-1}) = X_n - \bar{X}_{n-1}$, which gives

$$(n-1)s_n^2 - (n-2)s_{n-1}^2 = (X_n - \bar{X}_{n-1})^2 - n\left(\frac{X_n - \bar{X}_{n-1}}{n}\right)^2 = \frac{n-1}{n}(X_n - \bar{X}_{n-1})^2,$$

and finally

$$s_n^2 = \frac{n-2}{n-1}s_{n-1}^2 + \frac{1}{n}(X_n - \bar{X}_{n-1})^2.$$

---

5. (15 points) Let X be a realization from a hypergeometric distribution with parameters N, D, and n, that is, X is the number of defectives in a random sample (without replacement) of n objects from a population of size N objects which contains D defectives. Write down the probability distribution of X and express $f(x+1)$ as a function of $f(x)$ and the parameters of the distribution.

SOLUTION:

The hypergeometric probability distribution is given by

$$f(x) = \Pr(X = x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \qquad x = \max(0, n-N+D), \ldots, \min(n, D),$$

and if we divide $f(x+1)$ by $f(x)$ and do lots of canceling we get

$$f(x+1) = \frac{D-x}{x+1}\frac{n-x}{N-D-(n-x-1)}f(x).$$

---

6. (15 points) Let X be an $(n \times m)$ matrix and let $A = X^T X$.

a. How do we check whether X is of full rank?

b. How do we check whether A is positive definite if we don't know X?

c. How do we find the square root of A?

d. How do we find the determinant of A?

e. How do we solve a system of equations $Ax = b$?

SOLUTION:

a. X of full rank if and only if modified Gram Schmidt algorithm finishes normally.

b. A is positive definite if and only if the modified Cholesky decomposition finishes normally.

c. $A^{1/2} = LD^{1/2}$ where L and D are the factors in the MCD of A.

d. The determinant of A is the product of the diagonal elements of the matrix D in the MCD of A.

e. Solve $Ax = b$ by first finding the MCD of A and then solving two triangular and one diagonal system of equations.

---

7. (15 points) How would you modify the idea of Bresenham's line drawing method to draw a circle centered at pixel (c_0, r_0) of radius r pixels.

SOLUTION:

The basic ideas of the line drawing algorithm are that it uses recursion to determine which of two pixels to turn on at each step of the algorithm and that the decision rule is based on evaluating the function for a line at the midpoint between the two possible pixels at each step.

For a circle, one does the same thing. Now the function is $f(x, y) = (x - c_0)^2 + (y - r_0)^2 - r^2$ and a point is inside (outside) the circle if $f(x, y)$ is negative (positive). If we start at the leftmost point, and do the top semicircle, we have a steep section for one fourth of the semicircle, then a shallow section for half the semicircle, followed by another steep section for the last quarter of the semicircle. In a steep section we loop over rows and have to choose to go straight up one pixel or up one and over one. If the midpoint of the two choices is in the circle then we go straight up. We do something similar in the shallow portion.

The final thing to worry about is getting a recursion for the function at the midpoint that requires only integer arithmetic. This can be done but I really didn't expect you to do it.
