STAT 605:
Advanced Statistical Computations

David B. Dahl
Spring 2010

WRITTEN EXAMINATION

26 April 2010

Answer each question on a separate sheet of paper. Put your name and question number at the top of each sheet. You may use a basic (non-graphing) calculator.

1) {40 pts.} Consider the following integral:

$$\int_0^4 \frac{4^{(x-1)}}{(x+1)^3}dx.$$

(a) Using the composite trapezoidal rule with $M = 4$ equally-sized intervals of width $h = 1$, calculate the approximate value of this integral. Show your work.

(b) Write a function (in R or pseudo-code) to approximate this integral using Monte Carlo integration.

-0 2) {10 pts.} In the space below, provide the binary representation of a 32-bit IEEE-754 floating point number having the decimal value 20.375. Show your work on another sheet of paper. Under the boxes, label which bits are associated with the sign, mantissa, and exponent. Put your name of this page.

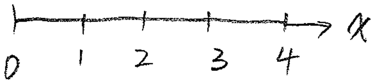
0	1	0	0	0	0	0	1	1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sign				exponent															mantissa												

3) {20 pts.}

Consider the 8×8 matrix to the right. You are tasked with testing the null hypothesis that the distribution of the ones is random versus the alternative hypothesis that the ones tend to be adjacent to each other. A pair of ones is adjacent if either: i) they are in the same row and are in adjacent columns, or ii) they are in the same column and are in adjacent rows. So the one in $(2, 3)$ is adjacent to the one in $(3, 3)$ but not the one in $(3, 2)$. Propose a nonparametric test statistic and describe how to compute its p-value.

	1	2	3					
1	0	1	1	0	0	0	1	0
2	0	0	1	0	1	0	1	1
3	0	1	1	0	1	0	0	0
	0	0	0	0	0	0	1	1
	0	0	0	0	0	1	1	0
	0	1	1	0	0	1	0	0
	0	0	0	0	0	1	1	1
	1	0	0	0	0	0	0	0

4) {10 pts.} Write a paragraph discussing the likely uses, benefits, and weaknesses of: (i) low-level programming languages (e.g. C, C++, Fortan, Java), (ii) high-level programming languages (e.g., R, MATLAB, etc.), and (iii) scripting languages (Ruby, Python, Perl, etc.) from the perspective of a research statistician.

1. a) ①  $h=1,\quad M=4$

$$\Rightarrow \int_0^4 f(x)\,dx \approx \frac{h}{2}\left(f(0)+f(4)\right) + h\sum_{i=1}^{3} f(x_i) \qquad ⊛$$

$$= \frac{1}{2}\left(f(0)+f(4)\right) + \left[f(1)+f(2)+f(3)\right]$$

where: $f(0)=\frac{1}{4}$, $f(1)=\frac{1}{8}$, $f(2)=\frac{4}{27}$

$f(3)=\frac{1}{4}$, $f(4)=\frac{64}{125}$

$$\Rightarrow \int_0^4 f(x)\,dx \approx \tfrac{1}{2}0.25 + 0.125 + 0.148 + 0.25 + \tfrac{1}{2}0.512$$

$$= 0.904$$

please see the back for proof of function ⊛ if needed.

b) ① $\int_0^4 \frac{4^{(x-1)}}{(x+1)^3}\,dx = \int_0^4 \frac{4^x}{(x+1)^3}\cdot\frac{1}{4}\,dx = E_f\left(\frac{4^x}{(x+1)^3}\right)$,

where f is the uniform distribution $(0,4)$ and $\begin{cases} f=\frac{1}{4} & \text{for } 0\le x\le 4 \\ f=0, & \text{otherwise}\end{cases}$

R:

```
nrep <- 50000
x <- numeric(nrep)
y <- numeric(nrep)

for (i in 1:nrep) {
x[i] <- runif(1, 0, 4)
y[i] <- 4^(x[i]) /(x[i]+1)^3
}
mean(y)
```

Here mean(y) will return the result of Monte-Carlo integration for $\int_0^4 \frac{4^{(x-1)}}{(x+1)^3}\,dx$

-0

(2) $20.375 = 20.375 = 20 + \frac{1}{4} + \frac{1}{8}$

$20 = 2^4 + 0(2^3) + 2^2 + 0(2^1) + 0(2^0)$

$= 1\ 0\ 1\ 0\ 0$

So $20.375 = 10100.011 = 1.0100011 \times 2^4$

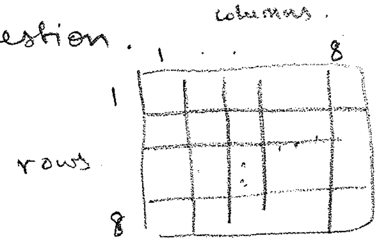
normalize exponent $\rightarrow 127 + 4 = 131$

$131 = 2^7 + 0(2^6) + 0(2^5) + 0(2^4) + 0(2^3) + 0(2^2) + 2^1 + 2^0$

$= 10000011$

Mantissa $= 0100011$

3. Let M denote the 8×8 matrix in question.



Here, H_0 : distribution of ones is random

H_1 : ones tend to be adjacent to each other

The following is how I would test the hypothesis.

step 1 I would use a permutation test. Let N = no. of permutations. Let $i = 1$

2 • Generate 2 sequences s_1 and s_2 where s_i = a sample of 8 values without replacement from 1 to 8, $i = 1, 2$.

• s_1 and s_2 basically tell us the way we shuffle the rows and columns respectively.

3. Compute the no. of adjacent pairs of ones, the test statistic, say $T[i]$

4. $i = i+1$. Go to step 2.

• Observe that T forms a null distribution.

• Let $\hat{T}$ = observed test statistic.

Then, the p-value for $\hat{T} = \dfrac{\#\{T_i s \geq T\}}{N}$.

-0

(4)

There are many things to consider when choo programming language such as: computing power, ease of programming implementation, functionality of program, etc. For doing many replications of the same calculation, low level languages might be the best choice since they can do calculations very quickly. This will not matter as much for computing one complex calculation, but when repeated many times (ie- resampling using Bootstrapping) a language more computationally efficient is preferable. A higher level language may be easier to program in, code is much more easily interpretable. Also, it might not be worth the time to learn a lower language if it won't save much computing time. Scripting languages are ideal for data cleaning and manipulation because they include many built in functions to accomplish these tasks. As a research statistician it may be useful to use any one, two, or even all three types to accomplish a project depending on what needs to be done. Scripting languages can do some basic statistical calculations and might be enough for some tasks. The type of program used will be project specific.

✓

-0