CMSC 460 - HW1

Gudjon Einar Magnusson

September 20, 2016

1

The following code constructs a bit array y by going through the series $2^{-1}, 2^{-2}, ..., 2^{-52}$ and subtracting 2^{-n} from x if $x > 2^{-n}$. $y_n = 1$ if $x > 2^{-n}$, otherwise it is zero.

```
function y = mydec2bin(x)
%x is a double precision value strictly less than 1.
%y is an integer array with 52 entries âĂę these are zero or one

y = zeros(1, 52);
leftover = x;

for n=1:52
    f = 2^-n;

    if (f <= leftover)
        leftover = leftover - f;
        y(n) = 1;
    end
end
```

2

2.1

x = 1; while 1+x > 1, x = x/2, pause(.02), end

Divides x by two and adds it to one until it underflows and the sum is

no longer greater than 1.
This prints 53 lines of output, the last two are 2^{-52} and 2^{-53}. 2^{-53} is too small to be added to 1.

x = 1; while x+x > x, x = 2*x, pause(.02), end
Multiplies x by two and adds it to itself until x overflows and the sum is no longer greater than x.
This prints 1024 lines of output, the last two are 2^{1023} and Inf.

x = 1; while x+x > x, x = x/2, pause(.02), end
Divides x by two and adds it to itself until x underflows and the sum is no longer greater than 1.
This prints 1075 lines of output, the last two are 2^{-1074} and 0. After gets below 2^{-1023} the number is no longer normalized.

2.2

2.2.1

$\mathcal{F}$ contains $2 \times (2^{11} - 2) \times 2^{52}$ elements. 2^{52} for every permutation of the mantissa, times 2^{11} for every permutation of the exponent except the two special values and times 2 for the sign.

2.2.2

2^{52} elements are in the range $[1 \leq x < 2]$, one for every permutation of the mantissa while the exponent is $0 + 1023$. The faction is $\frac{2^{52}}{|\mathcal{F}|}$

2.2.3

2^{52} elements are in the range $[\frac{1}{64} \leq x < \frac{1}{32}]$, one for every permutation of the mantissa while the exponent is $-6 + 1023$. The faction is $\frac{2^{52}}{|\mathcal{F}|}$

2.2.4

According to a random sample of a 100k numbers only about 85% satisfy the logical relation $x \times \frac{1}{x} = 1$. Many numbers that look like 1 are not actually

equal to 1.

```
%%Generate N random double presision numbers
N = 100000;

%% Random sign
s = -1 + ((rand(N, 1)>0.5) * 2);
%% Random exponent
e = floor(rand(N, 1) * 2^11);
%% Random mantissa
f = rand(N, 1) +1;

x = s .* 2.^(e-1023) .* f;
y = x.*(1./x) == 1;

sum(y) / N
```

2.3

The condition for the loop is that $s + t$ does not equal s. When t becomes small enough that $s + t = s$ the loop will end.

$x = \pi/2$
Accurate to within about 2.2×10^{-16}
Power series uses 11 terms.

$x = 11\pi/2$
Accurate to within about 2.1×10^{-10}
Power series uses 37 terms.

$x = 21\pi/2$
Accurate to within about 1.3×10^{-4} About 0.013%
Power series uses 60 terms.

$x = 31\pi/2$
Off by a factor of 3. Terribly inaccurate.
Power series uses 78 terms.

Power series can be used to accurately approximate functions on a small interval. The error increases quickly when the interval grows and more terms are needed. For a repeating function like $sin(x)$ this can fixed normalizing the value to a fixed range $[0, 2\pi]$.