/

Name: Zhao Yuhao *Student ID:* 9998 78342

1	25
2	32
3	31
T	88

STA 410/2102 — First Test — 2014-10-16

For all questions, show enough of your work to indicate how you obtained your answer. No books, notes, or calculators are allowed. You have 110 minutes to write this test. The total number of marks for all questions is 100 for undergrads in STA 410, and 105 for grad students in STA 2102.

Question 1: For the questions (a), (b), and (c) below, assume that numbers are represented in *decimal* (base ten) floating-point, with three significant decimal digits in the mantissa, and a range of -100 to $+100$ for the exponent. That is, the floating-point numbers are of the form $\pm 0.d_1d_2d_3 \times 10^E$, where d_1, d_2, and d_3 are digits from 0 to 9, and E is an integer between -100 and $+100$. Assume that floating-point arithmetic operations produce results that are properly rounded to the nearest representable number.

a) [9 marks] Find the result of each of the floating-point arithmetic operations below, writing your answers in the form $0.d_1d_2d_3 \times 10^E$, with d_1 not 0.

$(0.923 \times 10^5) + (0.900 \times 10^4)$ $= 0.101 \times 10^6$ ✓

$(0.200 \times 10^{-20}) * (0.103 \times 10^{10})$ $= 0.206 \times 10^{-11}$ ✓

$(0.188 \times 10^{12}) - (0.891 \times 10^{11})$ $= 0.989 \times 10^{11}$ ✓

0.1880
0.0891
0.0989

b) [10 marks] For each of the following true statements about arithmetic on real numbers, give a specific example showing that it is not necessarily true for computer arithmetic on floating-point numbers. In your examples, assume that overflow and underflow do not occur.

$2x - x = x.$ $x = 0.501 \times 10^0$ $2x = 0.100 \times 10^1$
$2x - x = 0.100 \times 10^1 - 0.501 \times 10^0 = 0.499 \times 10^0$
$\neq x$ ✓

$(x + x) + x = 3x.$ $x = 0.502 \times 10^0$
$(x+x) = 0.100 \times 10^1$ $3x = 0.151 \times 10^1$
$(x+x)+x = 0.150 \times 10^1$ $(x+x)+x \neq 3x$ ✓

c) [6 marks] Give an example showing it is possible for computation of $x * (y / z)$ to produce overflow (ie, for the result of some operation to require a value for E outside the range -100 to $+100$), but for computation of $(x * y) / z$ to not overflow. Write down *specific* values for x, y, and z, give the result of the second computation, and show that overflows occurs in the first computation.

$x = 0.1 \times 10^{-99}$ $y = 0.1 \times 10^{+99}$ $z = 0.1 \times 10^{-10}$

i) $\frac{y}{z} = \inf$ so $x * (\frac{y}{z}) = \inf$ overflow

ii) $(x * y) = 0.1 \times 10^{-1}$ $\frac{(x*y)}{z} = \frac{0.1 \times 10^{-1}}{0.1 \times 10^{-10}} = 0.1 \times 10^{9}$

thus $x * (\frac{y}{z})$ produce overflow $\frac{(x*y)}{z}$ produce normal value

Question 2: Suppose we have n i.i.d. (independent, identically-distributed) data points $x_1, \ldots, x_n$, that are real values in the interval $(-1, +1)$. We model these observations as having the distribution on $(-1, +1)$ with the following density function:

$$f(x) = (1 + \theta x)/2$$

where θ is an unknown model parameter in the interval $(-1, +1)$. We wish to find the maximum likelihood estimate for θ.

a) [10 marks] Write down the likelihood function, $L(\theta)$, and the log likelihood function, $\ell(\theta)$, based on the observations $x_1, \ldots, x_n$.

$$L(\theta) = \prod_{i=1}^{n} \frac{1+\theta x_i}{2} = \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} (1+\theta x_i)$$

$$\ell(\theta) = \sum_{i=1}^{n} \log \tfrac{1}{2}(1+\theta x_i)$$

$$\frac{1}{2^n}$$

b) [10 marks] Fill in the body of the following R function so that it will compute the first derivative of the log of the likelihood for theta based on the observations in the vector x:

$$\ell' = \sum \frac{\cancel{\theta}\, x_i}{1+\theta x_i}$$

```
log_lik_deriv1 <- function (theta, x) {
  n = length(x)
  res = 0
  for (i = 1:n) { res = res + theta x[i] /(1+(theta*x[i]))
  }
  res

}
```

4

You should use only basic R facilities, not deriv or D.

c) [10 marks] Fill in the body of the following R function so that it will compute the second derivative of the log of the likelihood for theta based on the observations in the vector x:

$$\ell'' = \sum - \frac{\theta^2 \, x_i}{(1+\theta x_i)^2}$$

```
log_lik_deriv2 <- function (theta, x) {
  n = length (x)
  res = 0
  for ( i = 1:n) {
    res = res - (theta^2)/((1+theta*x[i])^2)
  res

}
```

4

You should use only basic R facilities, not deriv or D.

d) [10 marks] Fill in the body of the following R function so that it returns the maximum likelihood estimate for θ given the data vector `x`. You should find the MLE using Newton iteration for `niters` iterations starting from `initial_theta`:

```
mle <- function (x, initial_theta, niters) {
  ~~n=0~~
  theta= initial-theta
  for (i=1: niters) {
    theta = theta - log-lik-deriv1 (theta,x) /log-like-deriv2(theta,x)
  }
  theta
}
```

You should use only basic R facilities and the two functions from parts (b) and (c) above, not the `nlm` or `optim` functions.

You do not need to do anything in this function to guard against the possibility of moving to a point outside the interval $(-1, +1)$. (We'll assume that the initial value is good enough to avoid this happening.)

Question 4: Suppose that n pairs of binary (0/1) values are produced, with the values in pair i being labelled as y_{i1} and y_{i2}. So the entire set of values produced is $y_{11}, y_{12}, y_{21}, y_{22}, \ldots, y_{n1}, y_{n2}$. We model these values as all being independent (both within and between pairs), with all values having probability p of being 1 and probability $1-p$ of being 0, where p is an unknown model parameter in $(0, 1)$.

If we observe all the y_{i1} and y_{i2} values, the log likelihood function will be

$$\begin{aligned} \ell(p) &= \log\Big[\prod_{i=1}^{n} p^{y_{i1}}(1-p)^{1-y_{i1}}\, p^{y_{i2}}(1-p)^{1-y_{i2}}\Big] \\ &= C\log(p) \;+\; (2n-C)\log(1-p) \end{aligned}$$

where $C = \sum_{i=1}^{n}(y_{i1} + y_{i2})$.

Note that the log likelihood function above depends only on n (which is known) and C, not on other aspects of the data.

We can find the MLE for p by differentiating $\ell(p)$ and setting it to zero:

$$\ell'(p) \;=\; \frac{C}{p} - \frac{2n-C}{1-p} \;=\; 0$$

from which we can see that the MLE is $\hat{p} = C\,/\,(2n)$.

Suppose, however, that for some pairs (chosen randomly without reference to the values), we do not observe the values in that pair, but only whether the values in the pair are the *same* or *different*.

For example, we might have data like the following:

i	y_{i1}	y_{i2}
1	0	1
2	1	1
3	same	
4	1	0
5	different	
6	0	1
7	same	

Let R be the sum of $y_{i1} + y_{i2}$ for all i where we know the exact values of y_{i1} and y_{i2}. Let S be the number of pairs where we know that $y_{i1} = y_{i2}$. Let D be the number of pairs where we know that $y_{i1} \neq y_{i2}$. For the example above, $R = 5$, $S = 2$, and $D = 1$.

We would like to find the MLE for p based on data like this (summarized by n, R, S, and D), using the EM algorithm.

a) [15 marks] In the E step of EM, we need to find the distribution for the missing data given the observed data and the current guess at the parameter, $p^{(t)}$. Since the log likelihood is a linear function of C, it will turn out that all we really need from this distribution is the expected value of C given the observed values for R, S, and D (along with n) and given the current guess for the parameter, $p^{(t)}$. Find a simple expression for this expected value.

i) for the pair same

$$y_{i1}+y_{i2} = \begin{cases} 0 & P = \frac{(1-p)^2}{p^2+(1-p)^2} = \frac{(1-p^{(t)})^2}{1+2p^{(t)2}-2p^{(t)}} \\ 2 & P = \frac{p^2}{1-2p} = \frac{(p^{(t)})^2}{1+2p^{(t)2}-2p^{(t)}} \end{cases}$$

$\therefore E(y_{i1}+y_{i2} \mid \text{same}) = 2 \times \frac{(p^{(t)})^2}{1+2(p^{(t)})^2-2p^{(t)}}$

ii) for the pair different

$y_{i1}+y_{i2} = 1 \quad \therefore E(y_{i1}+y_{i2} \mid \text{different}) = 1$

iii) for the pair we know values of y_{i1}, y_{i2} So we know $y_{i1}+y_{i2}$!

$$y_{i1}+y_{i2} = \begin{cases} 0 & P = (1-p^{(t)})^2 \\ 1 & P = 2p^{(t)}(1-p^{(t)}) \\ 2 & P = (p^{(t)})^2 \end{cases}$$

$E(y_{i1}+y_{i2} \mid \text{known } y_{i1}, y_{i2}) = 2p^{(t)}(1-p^{(t)}) + 2(p^{(t)})^2 = 2p^{(t)}$

$\therefore E(C) = 2 \times \frac{(p^{(t)})^2}{1+2(p^{(t)})^2-2p^{(t)}} \times S + 1 \times D + 2(n-S-D)\,p^{(t)}$

R

-4

b) [20 marks] In the M step of EM, we need find $p^{(t+1)}$ by maximizing the expected value of the log likelihood based on all y_{i1} and y_{i2} (as written above), with the expectation taken with respect to the distribution found in the E step. Find a simple expression for the expected log likelihood, and find a simple expression for the value of p that maximizes it.

Since $l(p^{(t)}) = C\log(p) + (2n-C)\log(1-p)$

$E(l(p^{(t)})) = E(C)\log(p) + (2n-E(C))\log(1-p)$

from a) we know

$$E(C) = \frac{2(p^{(t)})^2 S}{1+2(p^{(t)})^2-2p^{(t)}} + D + 2(n-S-D)p^{(t)} =$$

thus we have. $E(l(p^{(t)})) = \left(\frac{2(p^{(t)})^2 S}{1+2(p^{(t)})^2-2p^{(t)}} + D + 2(n-S-D)p^{(t)}\right)\log(p)$

$+ \left(2n - \frac{2(p^{(t)})^2 S}{1+2(p^{(t)})^2-2p^{(t)}} - D - 2(n-S-D)p^{(t)}\right)\log(1-p)$

take derivative and set to 0 we have:

$$p^{(t+1)} = \frac{E(C)}{2n} = \frac{1}{2n}\left(\frac{2(p^{(t)})^2 S}{1+2(p^{(t)})^2-2p^{(t)}} + D + 2(n-D-S)p^{(t)}\right)$$

✓

c) [5 marks, required for graduate students in STA 2102, undergrads in STA 410 can do it for bonus marks] Discuss whether the EM algorithm for this problem is guaranteed to converge to the global maximum of the likelihood, from any initial value for p in $(0, 1)$, or whether there may (for some data sets) be more than one local maximum of the likelihood.

let $f(p) = \frac{1}{2n}\left(\frac{2p^2 S}{1+2p^2-2p} + D + 2(n-D-S)p\right)$, $p \in (0,1)$

where p is the initial value for p

$\lim_{p \to 0} f(p) \doteq \frac{1}{2n}(1+D)$ which may not be the max value.

Thus the algorithm is not guaranteed to converge for initial value p closed to 0.

OK