Homework 3: Part II

Louis Bensard

February 27, 2018

(d)

```
p = 3
n = 200
m = p*(p+1)/2
maxit = 1000

mu = c(-1,1,2)
Sigma = matrix(c(1,0.7,0.7,0.7,1,0.7,0.7,0.7,1),3)
theta = vectorize(mu, Sigma,p)

#data
x = gen(n, p, mu, Sigma)

mu_0 = c(-1.5,1.5,2.3)
Sigma_0 = matrix(c(1,0.5,0.5,0.5,1,0.5,0.5,0.5,1),3)
theta_0 = vectorize(mu_0, Sigma_0, p)

tolgrad = 1e-9 ; tolerr = 1e-6
```

I - Steepest Ascent method with step halving

```
max_seeker(x, mu_0, Sigma_0, n, maxit, TRUE, FALSE, FALSE, tolerr, tolgrad)
```

```
##
## Iteration   Halving    log-likelihood    ||gradient||
##     1                      -853.5459        1.0e+03
##     1          1                  NA        NA
##     1          2                  NA        NA
##     1          3                  NA        NA
##     1          4                  NA        NA
##     1          5                  NA        NA
##     1          6                  NA        NA
##     1          7                  NA        NA
##     1          8                  NA        NA
##     1          9                  NA        NA
## ----------------------------------------------------------
##     2                      -783.9008        1.5e+03
##     2          1                  NA        NA
##     2          2                  NA        NA
##     2          3                  NA        NA
##     2          4                  NA        NA
##     2          5                  NA        NA
##     2          6                  NA        NA
##     2          7                  NA        NA
```

```
##     2           8                    NA         NA
##     2           9           -850.7135        1.2e+02
##     2           10           -808.6329        1.3e+02
##     2           11           -777.6896        1.5e+02
## ------------------------------------------------------------
## ...
## ...
## ...
##   998                       -709.9636        1.7e-05
##   998           1           -709.9636        1.0e-02
##   998           2           -709.9636        5.2e-03
##   998           3           -709.9636        2.6e-03
##   998           4           -709.9636        1.3e-03
##   998           5           -709.9636        6.4e-04
##   998           6           -709.9636        3.2e-04
##   998           7           -709.9636        1.5e-04
##   998           8           -709.9636        7.2e-05
##   998           9           -709.9636        3.1e-05
##   998           10           -709.9636        1.1e-05
##   998           11           -709.9636        8.6e-07
##   998           12           -709.9636        4.2e-06
##   998           13           -709.9636        6.8e-06
##   998           14           -709.9636        8.1e-06
##   998           15           -709.9636        8.7e-06
##   998           16           -709.9636        9.0e-06
## ------------------------------------------------------------
##   999                       -709.9636        1.6e-05
##   999           1           -709.9636        1.0e-02
##   999           2           -709.9636        5.0e-03
##   999           3           -709.9636        2.5e-03
##   999           4           -709.9636        1.3e-03
##   999           5           -709.9636        6.2e-04
##   999           6           -709.9636        3.1e-04
##   999           7           -709.9636        1.5e-04
##   999           8           -709.9636        7.0e-05
##   999           9           -709.9636        3.0e-05
##   999           10           -709.9636        1.1e-05
##   999           11           -709.9636        8.3e-07
##   999           12           -709.9636        4.1e-06
##   999           13           -709.9636        6.6e-06
##   999           14           -709.9636        7.8e-06
##   999           15           -709.9636        8.4e-06
##   999           16           -709.9636        8.7e-06
##   999           17           -709.9636        8.9e-06
##   999           18           -709.9636        8.9e-06
##   999           19           -709.9636        9.0e-06
## ------------------------------------------------------------
##  1000                       -709.9636        1.6e-05
##  1000           1           -709.9636        1.0e-02
##  1000           2           -709.9636        5.0e-03
##  1000           3           -709.9636        2.5e-03
##  1000           4           -709.9636        1.2e-03
##  1000           5           -709.9636        6.2e-04
##  1000           6           -709.9636        3.1e-04
```

```
##   1000          7             -709.9636        1.5e-04
##   1000          8             -709.9636        7.0e-05
##   1000          9             -709.9636        3.0e-05
##   1000          10             -709.9636        1.1e-05
##   1000          11             -709.9636        8.3e-07
##   1000          12             -709.9636        4.1e-06
##   1000          13             -709.9636        6.5e-06
##   1000          14             -709.9636        7.8e-06
##   1000          15             -709.9636        8.4e-06
##   1000          16             -709.9636        8.7e-06
##   1000          17             -709.9636        8.8e-06
##   1000          18             -709.9636        8.9e-06
## ----------------------------------------------------------
## ----------------------------------------------------------
```

After 1000 iterations, the condition on the norm of the gradient (<1e-9) hasn't been met, therefore the algorithm stops at maxit = 1000. If we would increase tolgrad to say 1e-5, the algorithm would stop after 323 iterations. But for our case, the maximum -709.9636 of the log-likelihood function is reached after 1000 iterations.

II - Fisher-scoring method with step-halving

```r
max_seeker(x, mu_0, Sigma_0, n, maxit, FALSE, TRUE, FALSE, tolerr, tolgrad)
```

```
##
## Iteration   Halving    log-likelihood    ||gradient||
##     1                       -853.5459        1.0e+03
##     1          0            -745.4340        8.3e+01
## ----------------------------------------------------------
##     2                       -745.4340        1.6e+02
##     2          0            -709.9636        1.2e-12
## ----------------------------------------------------------
## ...
## ...
## ...
##     1                       -853.5459        1.0e+03
##     1          0            -745.4340        8.3e+01
## ----------------------------------------------------------
##     2                       -745.4340        1.6e+02
##     2          0            -709.9636        1.2e-12
## ----------------------------------------------------------
##     3                       -709.9636        2.5e-12
##     3          0            -709.9636        2.0e-13
## ----------------------------------------------------------
## ----------------------------------------------------------
```

Both the norm of the gradient and the Modified Relative Error meet the criteria to stop the algorithm after 3 iterations. The maximum -709.9636 of the log-likelihood function is reached after 3 iterations.

III - Newton's method with step-halving

```r
max_seeker(x, mu_0, Sigma_0, n, maxit, FALSE, FALSE, TRUE, tolerr, tolgrad)
```

```
##
## Iteration   Halving    log-likelihood    ||gradient||
##     1                       -853.5459       1.0e+03
##     1           1                  NA       NA
##     1           2                  NA       NA
##     1           3                  NA       NA
##     1           4                  NA       NA
## ---------------------------------------------------------
##     2                       -848.6825       4.4e+03
##     2           1                  NA       NA
## ---------------------------------------------------------
## ...
## ...
## ...
##     7                       -709.9694       1.1e+01
##     7           0           -709.9636       9.9e-02
## ---------------------------------------------------------
##     8                       -709.9636       2.1e-01
##     8           0           -709.9636       4.3e-05
## ---------------------------------------------------------
##     9                       -709.9636       9.3e-05
##     9           0           -709.9636       6.9e-12
## ---------------------------------------------------------
## ---------------------------------------------------------
```

Both the norm of the gradient and the Modified Relative Error meet the criteria to stop the algorithm after 9 iterations. The maximum -709.9636 of the log-likelihood function is reached after 9 iterations.

(e)

```r
mu = c(-1,1,2)
Sigma_e = matrix(c(1,0.9,0.9,0.9,1,0.9,0.9,0.9,1),3)

#data
x_e = gen(n, p, mu, Sigma_e)

mu_0 = c(-1.5,1.5,2.3)
Sigma_0 = matrix(c(1,0.5,0.5,0.5,1,0.5,0.5,0.5,1),3)
theta_0 = vectorize(mu_0, Sigma_0, p)

max_seeker(x_e, mu_0, Sigma_0, n, maxit, TRUE, FALSE, FALSE, tolerr, tolgrad)
```

```
##
## Iteration   Halving    log-likelihood    ||gradient||
##     1                       -791.1402       9.4e+02
##     1           1                  NA       NA
##     1           2                  NA       NA
##     1           3                  NA       NA
##     1           4                  NA       NA
##     1           5                  NA       NA
##     1           6                  NA       NA
##     1           7                  NA       NA
##     1           8                  NA       NA
```

```
##     1          9                  NA        NA
##     1          10                  NA        NA
## ----------------------------------------------------------
##     2                      -670.8129        9.4e+02
##     2          1                  NA        NA
##     2          2                  NA        NA
##     2          3                  NA        NA
##     2          4                  NA        NA
##     2          5                  NA        NA
##     2          6                  NA        NA
##     2          7                  NA        NA
##     2          8                  NA        NA
##     2          9                  NA        NA
##     2          10                  NA        NA
##     2          11                  NA        NA
## ----------------------------------------------------------
## ...
## ...
## ...
##   998                      -505.6665        5.0e+00
##   998          1                  NA        NA
##   998          2                  NA        NA
##   998          3                  NA        NA
##   998          4           -564.1572        1.0e+03
##   998          5           -529.0145        4.7e+02
##   998          6           -514.1718        3.5e+02
##   998          7           -508.3743        2.5e+02
##   998          8           -506.4386        1.5e+02
##   998          9           -505.8697        8.3e+01
##   998          10           -505.7164        4.3e+01
##   998          11           -505.6777        2.1e+01
##   998          12           -505.6685        9.5e+00
##   998          13           -505.6666        3.6e+00
##   998          14           -505.6663        8.1e-01
## ----------------------------------------------------------
##   999                      -505.6663        1.4e+00
##   999          1                  NA        NA
##   999          2                  NA        NA
##   999          3                  NA        NA
##   999          4           -535.1274        2.7e+03
##   999          5           -510.1268        6.1e+02
##   999          6           -506.5793        2.3e+02
##   999          7           -505.8728        9.9e+01
##   999          8           -505.7145        4.6e+01
##   999          9           -505.6775        2.2e+01
##   999          10           -505.6687        1.1e+01
##   999          11           -505.6668        5.0e+00
##   999          12           -505.6663        2.3e+00
##   999          13           -505.6663        1.0e+00
## ----------------------------------------------------------
##  1000                      -505.6663        2.3e+00
##  1000          1                  NA        NA
##  1000          2           -594.1175        1.9e+03
##  1000          3           -539.8014        5.8e+02
```

```
## 1000          4           -518.7646       3.9e+02
## 1000          5           -510.0778       2.9e+02
## 1000          6           -506.9839       1.9e+02
## 1000          7           -506.0276       1.1e+02
## 1000          8           -505.7598       5.8e+01
## 1000          9           -505.6894       3.0e+01
## 1000          10           -505.6716       1.5e+01
## 1000          11           -505.6674       7.2e+00
## 1000          12           -505.6664       3.2e+00
## 1000          13           -505.6663       1.3e+00
## -----------------------------------------------------------
## -----------------------------------------------------------
```

After 1000 iterations, the condition on the norm of the gradient (<1e-9) hasn't been met, therefore the algorithm stops at maxit = 1000. The maximum -505.6663 of the log-likelihood function is reached after 1000 iterations.

But after 1000 iterations, in part (d) the norm of the gradient was 8.9e-06 as opposed to 1.3 for this part (e) so we can argue that part (d) is faster than part (e). The following proves more formally that this is correct:

```
theta_e = vectorize(mu, Sigma_e,p)

ev1 = eigen(-hessian(x, theta))[[1]]
ev2 = eigen(-hessian(x_e, theta_e))[[1]]

cr1 = ((max(ev1) - min(ev1))/(max(ev1) + min(ev1)))^2
cr2 = ((max(ev2) - min(ev2))/(max(ev2) + min(ev2)))^2

cat("cr1 =", cr1,"\ncr2 =", cr2)
```

```
## cr1 = 0.9508285
## cr2 = 0.9958086
```

The convergence ratio of part (d) is 0.9508, which is less than the convergence ratio of 0.9952 in part (e). Therefore, part (e) is indeed slower than part (d).

(f)

```
vect_ste = c()
mean_vect = c(mean(x[,1]), mean(x[,2]), mean(x[,3]))
Sig_cov = ((n-1)/n)*cov(x)
theta_mle = vectorize(mean_vect, Sig_cov,p)
f = fisher(x, theta_mle)
f_inv = solve(f)

for(i in 1:(m+p)){

    for(j in 1:(m+p)){

        if(i==j) vect_ste = c(vect_ste, sqrt(f_inv[i,j]))
    }
}

unvectorize(vect_ste, p)
```

```
## $mu
## [1] 0.07197462 0.07485894 0.07365860
##
## $Sigma
##             [,1]       [,2]       [,3]
## [1,] 0.10360692 0.09276979 0.09079161
## [2,] 0.09276979 0.11207721 0.09807797
## [3,] 0.09079161 0.09807797 0.10851179
```

\$mu is the standard error of the MLE $\hat{\mu}$ of μ. \$Sigma is the standard error of the MLE $\hat{\Sigma}$ of Σ.