2022년 2학기

Computational Statistics

HW#5

222STG10

김희숙

Problem

1. Missimg – EM method

1) Missing이 하나의 블럭

Data

w1 <- c(8,11,16,18,6,4,20,25,9,13)

w2 <- c(10,14,16,15,20,4,18,22,NA,NA)

initial values

mu1	mu2	sig11	sig22	sig12
13	14.59142	44.66667	26.29613	23.30100

Estimate, niter = 16

mu1	mu2	sig11	sig22	sig12
13	14.6423	44.66667	23.47737	20.78771

2) Missing이 두개의 블럭

Data

	X1	X2
1	1.55325851	1.91372035
2	1.46123101	-1.12739293
3	-1.14307818	-1.66216954
4	-0.64101971	1.08279750
5	1.58667859	3.07061416
6	-0.88092826	-0.16424174
7	1.21403860	-1.10030038
8	1.38171844	0.92975073
9	0.93376230	0.84668093
10	-2.09524714	-2.21228281
11	5.19289939	-0.05038494
12	3.52147495	2.76682971
13	2.34790897	2.61481685
14	1.57909332	1.10631296
15	1.15695818	0.66985132
16	-0.49513898	0.38722687
17	-3.99046442	-2.74345079
18	1.21882118	0.59718255
19	1.67749165	1.26853130
20	1.01824344	1.77187764
21	0.01368348	0.57353079
22	1.23259113	1.23747902
23	0.95951094	-1.46563015
24	-0.25025929	1.59583682
25	2.46246911	1.54286753
26	1.46891858	1.46251404
27	-0.12672791	2.56306823
28	-0.47870590	0.81201563
29	0.96121387	0.99127686
30	0.09041306	2.08651063
31	1.66960164	2.09354706
32	2.19532090	1.07586502
33	1.30304860	1.67268393
34	3.30174624	1.13638818
35	2.77996571	1.97939237
36	-1.15482877	-1.14602284
37	-0.48342226	0.12294514
38	0.86933786	0.31475389
39	1.36909889	0.59776465
40	3.79189231	2.51217732
41	0.70342151	0.11424224
42	1.07901409	1.46462659
43	3.02088069	3.19354872
44	-0.60633721	0.25950983
45	-0.13004498	0.11177657
46	0.10196162	0.02124661
47	1.11814700	1.82542753
48	1.85566550	-0.07753108
49	-0.32502178	1.44559419
50	0.81779917	0.34861239
51	1.98336136	2.61313005
52	2.10338436	-1.48634755
53	2.36763813	3.35270209
54	2.00724814	1.58361128
55	-2.18197333	-1.03498640
56	0.04287273	-0.08852180
57	1.84175447	1.97891581
58	1.86853804	0.74674803
59	1.27397067	0.88592022
60	1.13417337	0.52984906
71	2.18834156	0.71905888
72	0.54214905	-0.35698425
73	1.08936744	2.10341941
74	0.85429811	1.88145517
75	1.04885961	0.89103596
76	-1.17689322	-0.84709814
77	0.07848015	3.52563952
78	-1.82854904	-0.93865523
79	-0.71161244	1.50793791
80	2.15758910	0.96682896
81	NA	0.29121539
82	NA	1.40775037
83	NA	1.13668048
84	NA	-0.04471143
85	NA	1.51410988
86	NA	0.49068674
87	NA	-1.42366502
88	NA	0.81411556
89	NA	0.29203707
90	NA	1.64668380
91	0.45336075	NA
92	1.87266101	NA
93	0.66984083	NA
94	0.13512281	NA
95	2.92830215	NA
96	-0.11338535	NA
97	-0.37538090	NA
98	2.34976407	NA
99	0.96796426	NA
100	-0.02160807	NA

true

```
71   2.18834156  0.71905888
72   0.54214905 -0.35698425
73   1.08936744  2.10341941
74   0.85429811  1.88145517
75   1.04885961  0.89103596
76  -1.17689322 -0.84709814
77   0.07848015  3.52563952
78  -1.82854904 -0.93865523
79  -0.71161244  1.50793791
80   2.15758910  0.96682896
81  -1.02133691  0.29121539
82   1.07950614  1.40775037
83   1.71670202  1.13668048
84  -0.44756387 -0.04471143
85  -0.67921219  1.51410988
86  -0.12354616  0.49068674
87  -1.33281068 -1.42366502
88   1.59743045  0.81411556
89  -0.10346210  0.29203707
90   1.34775921  1.64668380
91   0.45336075  1.80190196
92   1.87266101  1.55490184
93   0.66984083  3.76521407
94   0.13512281  0.42670413
95   2.92830215  1.14728865
96  -0.11338535  0.62152154
97  -0.37538090  1.26501888
98   2.34976407  3.24372371
99   0.96796426  2.38372972
100 -0.02160807  0.08978028
```

initial values

mu1	mu2	sig11	sig22	sig12
0.9329402	0.8020199	2.1034892	1.0653989	1.6512363

Estimate, niter = 11, Time difference of 0.01269197 secs

mu1	mu2	sig11	sig22	sig12
0.9197149	0.8001877	1.904359	1.498269	0.9501219

True – Estimate

```
81  -1.6182889  0.00000000
82  -0.2254924  0.00000000
83   0.5836014  0.00000000
84  -0.8314891  0.00000000
85  -2.0516582  0.00000000
86  -0.8469922  0.00000000
87  -0.8422775  0.00000000
88   0.6688832  0.00000000
89  -0.7009352  0.00000000
90  -0.1087581  0.00000000
91   0.0000000  1.23438749
92   0.0000000  0.27927061
93   0.0000000  3.08969344
94   0.0000000  0.01796482
95   0.0000000 -0.65502267
96   0.0000000  0.33676782
97   0.0000000  1.11097989
98   0.0000000  1.73005640
99   0.0000000  1.55946945
100  0.0000000 -0.24076292
```

괜찮게 예측한 것으로 보인다. 단 93번째 값은 오차가 3으로크다.

2. Cell structure – EM method

no=176, na=182, nb=60, nab=17

initial values

p	q	r
0.3333333	0.3333333	0.3333334

Estimate, niter = 12, Time difference of 0.001867056 secs

p	q	r
0.2925356	0.09728839	0.610176

3. Peppered moths – EM method

nc=85, ni=196, nt=341

initial values

pc	pi	pt
0.3333333	0.3333333	0.3333333

Estimate, niter = 9, Time difference of 0.001459122 secs

pc	pi	pt
0.07083691	0.2293103	0.7331846

Code appendix

```
##### 1

#(1)

w1 <- c(8,11,16,18,6,4,20,25,9,13)

w2 <- c(10,14,16,15,20,4,18,22,NA,NA)


#initial values

mu1 <- mean(w1)

sig11 <- var(w1)

for (i in c(9, 10)){ # missing이 w2 9, 10번째 존재

  w2[i]  <-  mean(w2,  na.rm=TRUE)+cov(w1,  w2,  use='pairwise.complete.obs')/var(w1)*(w1[i]-
mean(w1))

}


mu2 <- mean(w2)

sig22 <- var(w2)

sig12 <- cov(w1,w2)

est <- c(mu2, sig22, sig12)


thres = 10^(-10); maxiter=1000

error <- 1

niter <- 0


while(niter <= maxiter & error >= thres) {

  # E-step

  for (i in c(9, 10)){

    w2[i] = mu2+sig12/sig11*(w1[i]-mu1)

  }


  # M-step

  mu2 = mean(w2)

  sig12 = (sum(w1*w2) - sum(w1)*sum(w2)/10)/10

  sig22 = (sum(w2*w2) - sum(w2)*sum(w2)/10)/10


  est_1 <- est

  est <- c(mu2, sig22, sig12)

  error <- sum(abs(est_1-est))

  niter <- niter + 1

}

w2;mu1;mu2;sig11;sig22;sig12;niter;
```

```
#(2)
library(MASS)
missing_em <-function(df, thres = 10^(-10), maxiter=1000) {
  start_time <- Sys.time()
  n <- nrow(df)
  #initial values
  mu1 = mean(df[,1],na.rm=TRUE)
  mu2 = mean(df[,2],na.rm=TRUE)
  sig11 = var(df[,1],na.rm=TRUE)
  sig22 = var(df[,2],na.rm=TRUE)
  sig12 =cov(df[,1], df[,2], use='pairwise.complete.obs')
  est <- c(mu1, mu2, sig11, sig12, sig22)
  print(est)
  error <- 1
  niter <- 0

  while(niter <= maxiter & error >= thres) {
    # E-step
    for (i in 81:90){
      df[i,1] = mu1+sig12/sig22*(df[i,2]-mu2)
    }

    for (i in 91:100){
      df[i,2] = mu2+sig12/sig11*(df[i,1]-mu1)
    }

    # M-step
    old_mu1 = mu1
    old_mu2 = mu2
    mu1 = mean(df[,1])
    mu2 = mean(df[,2])
    sig11 = (sum(df[,1]*df[,1]) - sum(df[,1])*sum(df[,1])/n)/n
    sig12 = (sum(df[,1]*df[,2]) - sum(df[,1])*sum(df[,2])/n)/n
    sig22 = (sum(df[,2]*df[,2]) - sum(df[,2])*sum(df[,2])/n)/n

    est_1 <- est
    est <- c(mu1, mu2, sig11, sig12, sig22)
    error <- sum(abs(est_1-est))
    niter <- niter + 1
```

```
  }
  end_time <- Sys.time()
  lst = list(mu1 =mu1,
             mu2= mu2,
             sig11 =sig11,
             sig22 = sig22,
             sig12 = sig12,
             time = end_time-start_time,
             niter = niter)
  print(true-df)
  return(lst)

}
df1 = data.frame(mvrnorm(100,c(1,1), matrix(c(2,1,1,2), nrow=2, byrow=T)))
true = df1
df1[81:90,1] = NA
df1[91:100,2] = NA
missing_em(df1)

##### 2
cell_structure <- function(no, na, nb, nab, thres = 10^(-10), maxiter=1000) {
  start_time <- Sys.time()

  #initial values
  n <- no+na+nb+nab
  est <- c(1/3, 1/3) ; r <- 1-est[1]-est[2]  # initial p, q, r
  print(est)
  error <- 1
  niter <- 0

  while(niter <= maxiter & error >= thres) {
    # E-step
    p <- est[1] ; q <- est[2]
    naa <- na*p^2/(p^2+2*p*r)
    nao <- na*2*p*r/(p^2+2*p*r)
    nbb <- nb*q^2/(q^2+2*q*r)
    nbo <- nb*2*q*r/(q^2+2*q*r)

    # M-step
    est_1 <- est
    est <- c((naa+nao/2+nab/2)/n, (nbb+nbo/2+nab/2)/n)  # new p, q
```

```
    error <- sum(abs(est-est_1))
    niter <- niter + 1

  }
  end_time <- Sys.time()
  lst = list(p.mle = est[1],
             q.mle = est[2],
             r.mle = 1-est[1]-est[2],
             time = end_time-start_time,
             niter = niter)
  return(lst)
}

cell_structure(176,182,60,17)



##### 3
peppered_moths <- function(nc, ni, nt, thres = 10^(-10), maxiter=1000) {
  start_time <- Sys.time()
  #initial values
  n <- nc+ni+nt
  est <- c(1/3, 1/3, 1/3)  # initial pc, pi, pt
  print(est)

  error <- 1
  niter <- 0

  while(niter <= maxiter & error >= thres) {
    # E step
    pc <- est[1] ; pi <- est[2] ; pt <- 1-est[1]-est[2]
    ncc <- nc*pc^2/(pc^2+2*pc*pi+2*pc*pt)
    nci <- nc*2*pc*pi/(pc^2+2*pc*pi+2*pc*pt)
    nct <- nc*2*pc*pt/(pc^2+2*pc*pi+2*pc*pt)
    nii <- ni*pi^2/(pi^2+2*pi*pt)
    nit <- ni*2*pi*pt/(pi^2+2*pi*pt)

    # M step
    est_1 <- est
    est <- c((2*ncc+nci+nct)/(2*n), (2*nii+nit+nct)/(2*n), (2*nt+nct+nit)/(2*n)) # new pc, pi,
pt
```

```
    error <- sum(abs(est-est_1))
    niter <- niter + 1
  }
  end_time <- Sys.time()
  lst = list(estimatedPc = est[1],
             estimatedPi = est[2],
             estimatedPt = est[3],
             time = end_time-start_time,
             niter = niter)

  return(lst)
}

peppered_moths(85, 196, 341)
```