Project 8 Nested Effect Models

Team C - Minghang Li, Xiaocheng Yang, Xinyi Chen

April 18, 2024

Problem 20: Classical NEMs

Subproblem 1

Construct transitive closure and define Φ

Define a function to make any Φ matrix transitive closed by powering it up until convergence.

Construct Φ for Model (a).

```
phi_a <- array(dim = c(5, 5),
               dimnames = list(c("S1", "S2", "S3", "S4", "S5"),
                               c("S1", "S2", "S3", "S4", "S5")))
#                   S1, S2, S3, S4, S5
phi_a["S1",] <- c(1,  0,  1,  1,  0)
phi_a["S2",] <- c(0,  1,  0,  0,  1)
phi_a["S3",] <- c(0,  0,  1,  1,  1)
phi_a["S4",] <- c(0,  0,  0,  1,  1)
phi_a["S5",] <- c(0,  0,  0,  0,  1)

phi_a <- transitive.closure(phi_a)
phi_a
```

```
##    S1 S2 S3 S4 S5
## S1  1  0  1  1  1
## S2  0  1  0  0  1
## S3  0  0  1  1  1
## S4  0  0  0  1  1
## S5  0  0  0  0  1
```

Construct Φ for Model (b).

```
phi_b <- array(dim = c(5, 5),
               dimnames = list(c("S1", "S2", "S3", "S4", "S5"),
                               c("S1", "S2", "S3", "S4", "S5")))
#                   S1, S2, S3, S4, S5
phi_b["S1",] <- c(1,  0,  0,  1,  0)
phi_b["S2",] <- c(0,  1,  0,  0,  1)
phi_b["S3",] <- c(1,  0,  1,  1,  1)
phi_b["S4",] <- c(0,  0,  0,  1,  1)
phi_b["S5",] <- c(0,  0,  0,  0,  1)

phi_b <- transitive.closure(phi_b)
phi_b
```

```
##    S1 S2 S3 S4 S5
## S1  1  0  0  1  1
## S2  0  1  0  0  1
## S3  1  0  1  1  1
## S4  0  0  0  1  1
## S5  0  0  0  0  1
```

Define Θ

Define Θ for Model (a).

```
theta_a <- array(dim = c(5, 6),
                 dimnames = list(
                   c("S1", "S2", "S3", "S4", "S5"),
                   c("E1", "E2", "E3", "E4", "E5", "E6")
                 ))
#                    E1, E2, E3, E4, E5, E6
theta_a["S1",] <- c(0,  0,  0,  0,  0,  0)
theta_a["S2",] <- c(0,  0,  0,  1,  0,  1)
theta_a["S3",] <- c(1,  1,  0,  0,  0,  0)
theta_a["S4",] <- c(0,  0,  1,  0,  0,  0)
theta_a["S5",] <- c(0,  0,  0,  0,  1,  0)
theta_a
```

```
##    E1 E2 E3 E4 E5 E6
## S1  0  0  0  0  0  0
## S2  0  0  0  1  0  1
## S3  1  1  0  0  0  0
## S4  0  0  1  0  0  0
## S5  0  0  0  0  1  0
```

Define Θ for Model (b).

```
theta_b <- array(dim = c(5, 6),
                 dimnames = list(
                   c("S1", "S2", "S3", "S4", "S5"),
                   c("E1", "E2", "E3", "E4", "E5", "E6")
                 ))
#                    E1, E2, E3, E4, E5, E6
theta_b["S1",] <- c(1,  1,  0,  0,  0,  0)
theta_b["S2",] <- c(0,  0,  0,  1,  0,  1)
theta_b["S3",] <- c(0,  0,  0,  0,  0,  0)
theta_b["S4",] <- c(0,  0,  1,  0,  0,  0)
theta_b["S5",] <- c(0,  0,  0,  0,  1,  0)
theta_b
```

```
##    E1 E2 E3 E4 E5 E6
## S1  1  1  0  0  0  0
## S2  0  0  0  1  0  1
## S3  0  0  0  0  0  0
## S4  0  0  1  0  0  0
## S5  0  0  0  0  1  0
```

Determine the corresponding expected effect patterns (F)

```
F_a <- phi_a %*% theta_a
F_a
```

```
##    E1 E2 E3 E4 E5 E6
## S1  1  1  1  0  1  0
## S2  0  0  0  1  1  1
## S3  1  1  1  0  1  0
## S4  0  0  1  0  1  0
## S5  0  0  0  0  1  0
```

```
F_b <- phi_b %*% theta_b
F_b
```

```
##    E1 E2 E3 E4 E5 E6
## S1  1  1  1  0  1  0
## S2  0  0  0  1  1  1
## S3  1  1  1  0  1  0
## S4  0  0  1  0  1  0
## S5  0  0  0  0  1  0
```

Subproblem 2

If we assume no noise (no false positives and false negatives)… then the D matrix is simply the F matrix transpose.

```
D_a <- t(F_a)
D_a
```

```
##    S1 S2 S3 S4 S5
## E1  1  0  1  0  0
## E2  1  0  1  0  0
## E3  1  0  1  1  0
## E4  0  1  0  0  0
## E5  1  1  1  1  1
## E6  0  1  0  0  0
```

```
D_b <- t(F_b)
D_b
```

```
##    S1 S2 S3 S4 S5
## E1  1  0  1  0  0
## E2  1  0  1  0  0
## E3  1  0  1  1  0
## E4  0  1  0  0  0
## E5  1  1  1  1  1
## E6  0  1  0  0  0
```

Given the discrete data D_a and D_b (sorry for the different notation from the exercise pdf) it's not possible to tell apart the two models because they are identical.

```
all.equal(D_a, D_b)
```

```
## [1] TRUE
```

Subproblem 3

Calculate the marginal log-likelihood ratio (network score) given the data by setting the false positive rate to be 5% and the false negative rate to be 1%.

```
network_score_a <- scoreAdj(
  D_a,
  adj = phi_a,
  method = "disc",
  fpfn = c(0.05, 0.01),
  logtype = exp(1)
)$score
network_score_a
```

```
## [1] 41.79955
```

```
network_score_b <- scoreAdj(
  D_b,
  adj = phi_b,
  method = "disc",
  fpfn = c(0.05, 0.01),
  logtype = exp(1)
)$score
network_score_b
```

```
## [1] 41.79955
```

Problem 21: Hidden Markov NEMs

Subproblem 1

Compute the transition probabilities from $G_t = u$ to $G_{t+1} \in \{v_1, v_2\}$ for different smoothness parameter $\lambda \in \{0.1, \dots, 0.9\}$.

By definition, the probability of transition from network u to network v is calculated by:

$$\begin{aligned} T_{uv} &= P(\Phi_{t+1} = v | \Phi_t = u) \\ &= \frac{1}{C_u}(1-\lambda)^{s_{uv}} \cdot \lambda \end{aligned}$$

The distance s_{uv} is defined as

$$s_{uv} = ||u - v||_1 := \sum_i \sum_{i'} |u_{ii'} - v_{ii'}|$$

The normalizing constant C_u is defined as

$$C_u = \sum_w (1-\lambda)^{s_{uw}} \cdot \lambda$$

where w is all possible networks given the S genes at hand.

So the basic implementation idea would be:

1. Represent u, v_1 and v_2 using adjacency matrix
2. Compute s_{uv_1} and s_{uv_2} by "diff"ing the pairs of matrices respectively
3. Generate all the networks w using `mnem`, compute all the s_{uw}.

4. Compute transition probability for each λ

Implementation in R is in the following code blocks.

```
S_genes <- c("S1", "S2", "S3", "S4")
```

```
# Initialization of u
u <- array(dim = c(4, 4),
           dimnames = list(S_genes, S_genes))
#              S1, S2, S3, S4
u["S1",] <- c(1,  1,  1,  0)
u["S2",] <- c(0,  1,  1,  1)
u["S3",] <- c(0,  0,  1,  1)
u["S4",] <- c(0,  0,  0,  1)
u <- transitive.closure(u)
u
```

```
##    S1 S2 S3 S4
## S1  1  1  1  1
## S2  0  1  1  1
## S3  0  0  1  1
## S4  0  0  0  1
```

```
# Initialization of v1
v1 <- array(dim = c(4, 4),
            dimnames = list(S_genes, S_genes))
#               S1, S2, S3, S4
v1["S1",] <- c(1,  1,  1,  0)
v1["S2",] <- c(0,  1,  1,  1)
v1["S3",] <- c(0,  0,  1,  0)
v1["S4",] <- c(0,  0,  0,  1)
v1 <- transitive.closure(v1)
v1
```

```
##    S1 S2 S3 S4
## S1  1  1  1  1
## S2  0  1  1  1
## S3  0  0  1  0
## S4  0  0  0  1
```

```
# Initialization of v2
v2 <- array(dim = c(4, 4),
            dimnames = list(S_genes, S_genes))
#               S1, S2, S3, S4
v2["S1",] <- c(1,  0,  0,  0)
v2["S2",] <- c(1,  1,  1,  0)
v2["S3",] <- c(1,  0,  1,  0)
v2["S4",] <- c(1,  0,  0,  1)
v2 <- transitive.closure(v2)
v2
```

```
##    S1 S2 S3 S4
## S1  1  0  0  0
## S2  1  1  1  0
## S3  1  0  1  0
## S4  1  0  0  1
```

```r
# compute s_uv1 and s_uv2
s_uv1 <- sum(u != v1)
s_uv2 <- sum(u != v2)
```

```r
# generate all the possible networks
all_networks <- mnem:::enumerate.models(S_genes, trans.close = TRUE)
```

```
## Generated 355 unique models ( out of 4096 )
```

```r
# compute s_uw for all networks
num_cores <- detectCores()
registerDoParallel(num_cores)
start <- Sys.time()
s_uw <-
  foreach (
    i = 1:length(all_networks),
    .combine = c,
    .packages = c("foreach")
  ) %dopar% {
    sum(u != all_networks[[i]])
  }
end <- Sys.time()
end - start
stopImplicitCluster()
```

```r
compute_C <- function(lambda, s_uw) {
  res <-
    foreach (
      i = 1:length(s_uw),
      .combine = c,
      .packages = c("foreach")
    ) %dopar% {
      (1 - lambda) ^ s_uw[i] * lambda
    }
  return(sum(res))
}
```

```r
# Compute transitive probability
registerDoParallel(num_cores)
start <- Sys.time()
trans_prob <-
  foreach (
    lambda = seq(0.1, 0.9, by = 0.1),
    .combine = rbind,
    .packages = c("foreach")
  ) %dopar% {
    C_u <- compute_C(lambda = lambda, s_uw = s_uw)
    res1 <- (1 - lambda) ^ s_uv1 * lambda / C_u
    res2 <- (1 - lambda) ^ s_uv2 * lambda / C_u
    return(c(res1, res2))
  }
end <- Sys.time()
end - start
```

```
## Time difference of 0.1622491 secs
```

```r
stopImplicitCluster()
colnames(trans_prob) <- c("v1", "v2")
rownames(trans_prob) <- sprintf("%.1f", seq(0.1, 0.9, 0.1))
trans_prob
```

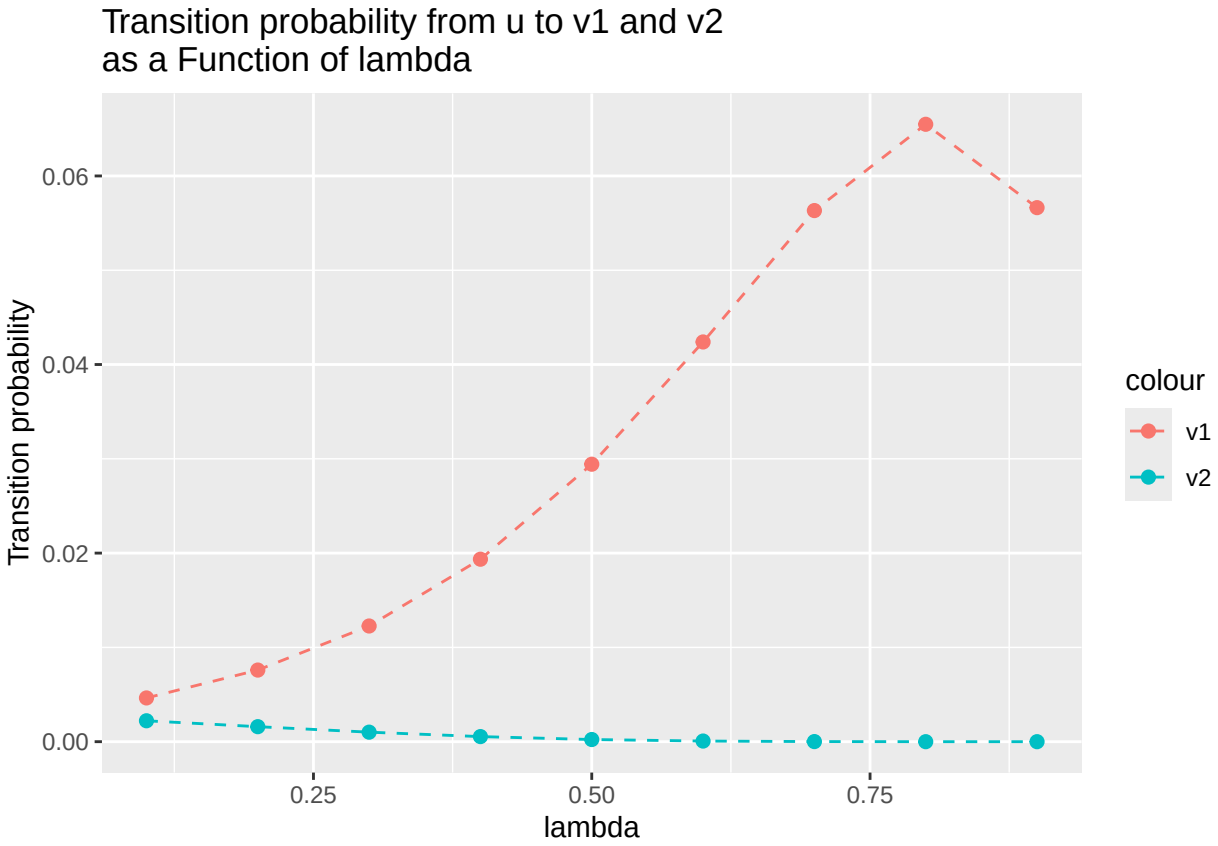
```
##              v1           v2
## 0.1 0.004640039 2.219316e-03
## 0.2 0.007597382 1.593287e-03
## 0.3 0.012270196 1.010503e-03
## 0.4 0.019353227 5.417665e-04
## 0.5 0.029428247 2.299082e-04
## 0.6 0.042393843 6.945807e-05
## 0.7 0.056328212 1.231898e-05
## 0.8 0.065478114 8.381199e-07
## 0.9 0.056637363 5.663736e-09
```

Subproblem 2

Plot the transition probabilites as a function of λ for v_1 and v_2.

```r
df <-
  data.frame(lambda = row.names(trans_prob),
             trans_prob,
             row.names = NULL)
df$lambda <- as.numeric(df$lambda)
```

```r
ggplot(data = df, aes(x = lambda, group = 1)) +
  geom_line(aes(y = v1, color = "v1"), linetype = "dashed") +
  geom_point(aes(y = v1, color = "v1"), size = 2) +
  geom_line(aes(y = v2, color = "v2"), linetype = "dashed") +
  geom_point(aes(y = v2, color = "v2"), size = 2) +
  xlab("lambda") +
  ylab("Transition probability") +
  ggtitle("Transition probability from u to v1 and v2\nas a Function of lambda")
```



We can see that the transition probabilities of v_1 and v_2 converge when λ is small and differ greatly when λ is large. Dissimilar networks get penalized and result in lower transition probability as λ increase. For similar networks like v_1, the probability to transit into them increases as we increase λ.

Problem 22: Mixture NEMs

Subproblem 1

Determine cellular perturbation map ρ where $\rho_{ic} = 1$ if cell c is perturbed by a knowdown of S-gene i.

```
S_genes <- c("S1", "S2")
E_genes <- c("E1", "E2")
cells   <- c("C1", "C2", "C3", "C4")
```

```
rho = array(dim = c(2, 4),
            dimnames = list(S_genes, cells))
#               C1, C2, C3, C4
rho["S1",] <- c(1,  0,  1,  0)
rho["S2",] <- c(0,  1,  1,  1)
rho
```

```
##    C1 C2 C3 C4
## S1  1  0  1  0
## S2  0  1  1  1
```

Subproblem 2

Assume $\{C_1, C_2\}$ are generated from F_1 and $\{C_3, C_4\}$ are generated from F_2, compute the *noiseless* log odds matrix R, where $R_{jc} > 0$ means that the perturbation on cell c has an effect on E-gene j.

(a) Compute expected effect pattern $(\rho^T \phi_k \theta_k)^T$

Of course need to define Φ and Θ again for F_1 and F_2.

```
phi_1 <- array(dim = c(2, 2), dimnames = list(S_genes, S_genes))
#                  S1, S2
phi_1["S1",] <- c(1,  1)
phi_1["S2",] <- c(0,  1)

theta_1 <- array(dim = c(2, 2), dimnames = list(S_genes, E_genes))
#                    E1, E2
theta_1["S1",] <- c(1,  0)
theta_1["S2",] <- c(0,  1)
```

```
phi_2 <- array(dim = c(2, 2), dimnames = list(S_genes, S_genes))
#                    S1, S2
phi_2["S1", ] <- c( 1,  0)
phi_2["S2", ] <- c( 1,  1)

theta_2 <- array(dim = c(2, 2), dimnames = list(S_genes, E_genes))
#                     E1, E2
theta_2["S1", ] <- c( 0,  1)
theta_2["S2", ] <- c( 1,  0)
```

```
EEP_1 <- t(t(rho) %*% phi_1 %*% theta_1)
EEP_1[EEP_1 > 1] <- 1
EEP_1
```

```
##    C1 C2 C3 C4
## E1  1  0  1  0
## E2  1  1  1  1
```

```
EEP_2 <- t(t(rho) %*% phi_2 %*% theta_2)
EEP_2[EEP_2 > 1] <- 1
EEP_2
```

```
##    C1 C2 C3 C4
## E1  0  1  1  1
## E2  1  1  1  1
```

(b) Extract the corresponding colum from the expected effect patterns and put it into R

```
# C1, C2 from F1; C3, C4 from F2
R <- cbind(EEP_1[, 1:2], EEP_2[, 3:4])
R[R == 0] <- -1
R
```

```
##    C1 C2 C3 C4
## E1  1 -1  1  1
## E2  1  1  1  1
```

Subproblem 3

Calculate the responsibilities Γ given mixture weights $\pi = (0.44, 0.56)$. Then update the mixture weights. (Well, EM again!)

```
pi <- c(0.44, 0.56)
```

```
# log likelihood
L1 <- t(EEP_1) %*% R
L2 <- t(EEP_2) %*% R

gamma <- rbind(diag(pi[1] * exp(L1) / (pi[1] * exp(L1) + pi[2] * exp(L2))),
               diag(pi[2] * exp(L2) / (pi[1] * exp(L1) + pi[2] * exp(L2))))
gamma
```

```
##             C1        C2   C3        C4
## [1,] 0.6811014 0.6811014 0.44 0.2242338
## [2,] 0.3188986 0.3188986 0.56 0.7757662
```

```
pi[1] <- mean(gamma[1,])
pi[2] <- mean(gamma[2,])
```

```
pi
```

```
## [1] 0.5066091 0.4933909
```