

Group Testing Synthetic Experiments

(1) Construct Mock Data Set

Objective: create a mock data set with a list of individuals, assigned to groups, with true infected status

Method: 1. Fix number of people per group (n) 2. Set probability per group (p = probability of infection)

3. Randomly select x people to infect in the population: The number of infected people in a group is now binomial $\text{bin}(n - x, p)$

```
# set parameters
g <- 4 # number of groups
n <- 5 # number of people per group
p <- c(.1, .2, .3, .4) # infection probability per group (could put a probability distribution, for now)
x <- 5 # number of people to infect in population

# construct dataframe
id <- seq(1, g*n, by = 1) # individual ID
group <- rep(1:g, each = n) # assign group numbers
prob_infect <- rep(p, each = n) # assign infection probabilities
infected <- rep(0, g*n) # infection status: 1 = infected; 0 = not infected
data <- as.data.frame(cbind(id, group, prob_infect, infected))

# randomly select individuals to infect
i <- sample(data$id, size = x)
data$infected[i] <- 1 # change infection status
```

(2) Apply Bilder's Algorithm to Mock Data

How many steps would it take for Bilder's algorithm to find the infected people in the mock data set?

- optimal retesting configuration (ORC): the configuration which minimizes $E(T)$ when ordered individuals are successively placed into subgroups.

Code from Bilder paper: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4495770/#SD1> (for computing pool sizes, per the algorithm described in the paper)

```
# http://www.chrisbilder.com/grouptesting/ (see paper 11)
#Program that contains R functions used to find ORC and CRC
source("functions.R")

#####
# Section 3 - Find ORC and CRC for  $p = 0.05$ ,  $\alpha = 0.1$ ,  $I = 5$ 
# See Table 2 in the on-line supplementary materials for the same numerical values

# Sensitivity and specificity
Se <- 0.95
Sp <- 0.95
```

```

#Order statistics for individual probabilities
p.a0.1 <- beta.dist(p = 0.05, a = 0.1, grp.sz = 5)
p.a0.1

## [1] 0.0001378157 0.0015697472 0.0098221369 0.0460956109 0.1923991296

#CRC
CRC.a0.1 <- get.CRC(p = p.a0.1, se = Se, sp = Sp, stages = 3)

## [1] "Three stage procedure"
CRC.a0.1$ET      # E(T)

## [1] 1.777426
CRC.a0.1$ET/5    # E(T)/I

## [1] 0.3554852
CRC.a0.1$I2      # Subgroup sizes for second stage

## [1] 4 1

#ORC
ORC.a0.1 <- get.CRC(p = p.a0.1, se = Se, sp = Sp, everycase = TRUE, stages = 3)

## [1] "Warning: if group size is large (>18) program may take excessive time"
## [1] "Three stage procedure"
## [1] "ORC is the optimal looking at every possible configuration"
ORC.a0.1$ET      # E(T)

## [1] 1.777426
ORC.a0.1$ET/5    # E(T)/I

## [1] 0.3554852
ORC.a0.1$I2      # Subgroup sizes for second stage

## [1] 4 1

```

Next steps: compute optimal group sizes, apply them to the mock data set, see how many tests are needed to detect the infectious people in the population (accounting for sensitivity and specificity of test?)