Exercise 1: What is a likelihood in general? What does it mean when an estimator is a maximum likelihood estimator?

Solution:
Likelihood is used to describe the joint probability of achieving our sampled data given certain model parameters. Said another way, we fix our sample x and we vary our model parameters θ the likelihood gives us the probability of achieving x for a certain θ. Generally this is written in the form of a function $P(x|\theta)$ and the maximum likelihood estimator involves selecting model parameters which maximize said function.

Exercise 2: We are doing a small mark-recapture study. First we catch 20 fish in a lake and tag each one. Then, a few days later we catch 40 fish and notice that 6 were tagged.

a. Is this direct or indirect sampling? Why?

   Solution:
   This method would be direct mark-recapture sampling. In an indirect scheme we would decide prior to the sample the number of units we would mark and recapture, then on the recapture survey we would continue to survey fish until we reach the desired number of recaptures. In this method the size of the recapture sample is random m, not the number of recaptures themselves.

b. Produce a 95 percent confidence interval for the population size (using either Lincoln-Petersen or Chapman estimator).

   Solution:
   Since the text describes the Chapman estimator as the one often used in practice, it is how we will proceed.
   Code:

```
> n = 20
> m = 40
> r = 6
> N = ((n+1)*(m+1))/(r+1) - 1
[1] 122
```

```
> Var = ((n+1)*(m+1)*(n-r)*(m-r))/(((r+1)^2)*(r+2))
[1] 1045.5
> CI95 <- c(N + 2*sqrt(Var), N - 2*sqrt(Var))
[1] 186.66838  57.33162
```

c. What assumptions do we need to compute the mark-recapture estimator in (b)?

Solution:
For direct mark recapture sampling estimators we assume that the second sample is indeed a SRS, without replacement, we assume that the population is closed, and we assume that our marked samples stayed mark for the second sample.

Exercise 3: Suppose we visited a site at five different equally- spaced times and trapped, marked and released animals each time. We will assume a closed model.

a. What is a closed model? When is it reasonable to assume that the population is closed?

Solution:
A closed model is one that does not allow immigration, emigrations, births, or deaths. It is reasonable to assume that a population is closed when the sampling sessions are closer together in time.

b. If the following data is what we obtained, look at the data (don't analyze it) and try to explain what might be happening (and which model, M_0, M_t, M_h, M_{th}, etc., would be appropriate).

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Solution:
Given that the last sample session has no recaptures, and the total number of recaptures at each session seems to decrease I think the M_t model, which allows the capture probability of each animal to change from session to session, would be appropriate.

Now, try to explain what type of closed population model might fit with this data:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Here I would say the probability of capture varies from animal to animal. We can see that there is large variation in the row sums of the matrix, while the column sum stay relatively that same. I would say the M_h model would be appropriate.

Exercise 4: Run the following code in R and choose which model (based on AIC) seems to have the best fit. What abundance does this model estimate?

Solution:
Fitting the capture data with the closedp() function, we get that the M_b model achieves the best fit with the lowest AIC. I would also consider the M_0 given its relatively similar score and greater simplicity,
Code:

```
> library(Rcapture)
> dat <- matrix(ncol=5, byrow=TRUE,
+   c(0,0,1,0,
+     0,0,0,1,
+     0,0,1,1,
+     0,0,0,1,
+     0,0,0,1,
+     1,0,0,1,
+     0,1,0,0,
+     0,0,0,1,
+     0,0,1,1,
+     0,1,0,1))

> out <- closedp(dat)
> out

Number of captured units: 8

Abundance estimations and model fits:
                abundance stderr deviance df    AIC    BIC infoFit
M0                    9.9    2.1   18.784 29 37.397 37.556      OK
Mt                    9.5    1.8   14.228 25 40.842 41.319      OK
Mh Chao (LB)          9.9    2.1   18.784 29 37.397 37.556      OK
Mh Poisson2           8.5    1.1   17.803 28 38.417 38.655      OK
Mh Darroch            8.3    0.8   17.850 28 38.463 38.702      OK
Mh Gamma3.5           8.1    0.6   17.874 28 38.488 38.726      OK
Mth Chao (LB)         9.5    1.8   14.228 25 40.842 41.319      OK
Mth Poisson2          8.5    1.1   13.451 24 42.064 42.621      OK
Mth Darroch           8.3    0.8   13.494 24 42.108 42.664      OK
Mth Gamma3.5          8.2    0.6   13.516 24 42.130 42.686      OK
Mb                    8.1    0.3   15.276 28 35.890 36.128      OK **
Mbh                   8.1    1.4   15.219 27 37.833 38.151      OK
```

Exercise 5: Consider the M_t model, where the probabilities are different for each time period. If there are 3 time periods, the parameters are p_1, p_2, p_3.
Suppose we have the dataset (one animal per line):

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Write down the likelihood for this data.

Solution:
First let's consider the probability under M_t for each pattern in the data,

$$\begin{bmatrix} & \text{Pattern} & & \text{Prob under } M_t \\ 1 & 0 & 0 & (p_1)(p_2-1)(p_3-1) \\ 1 & 0 & 0 & (p_1)(p_2-1)(p_3-1) \\ 0 & 1 & 0 & (p_1-1)(p_2)(p_3-1) \\ 0 & 1 & 1 & (p_1-1)(p_2)(p_3) \end{bmatrix}$$

Multiplying we get the following likelihood function,

$$L(p_1, p_2, p_3) = (p_1)^2(p_1-1)^2(p_2)^2(p_2-1)^2(p_3)(p_3-1)^3$$

Exercise 6: The following data will cause problems with the open population model unless we force equal capture probabilities across all times. Why do you think that is the case? (The last column is the number of times that pattern occurred, so that the pattern 1 0 0 0 occurred 7 times.)

Solution:
It is problematic to have only four sampling sessions and allow unequal capture probability mainly because we already have to throw out the first and last sessions because we cannot determine the difference between the capture probability and survival probability at the endpoints(if the last session was a zero, is it because the animal died or because we simply didn't recapture it). If we are going to allow unequal probability it is best if we do many, many recaptures.
When we let capture probabilities be constant, survival probabilities can be calculated, as is shown in the following code example,
Code:

```
> dat = matrix(ncol=5,byrow=TRUE,
+   c(1, 0, 0, 0, 7,
+     0, 1, 0, 0, 20,
+     0, 0, 1, 0, 37,
+     0, 0, 0, 1, 70,
+     1, 1, 0, 0, 7,
+     1, 0, 1, 0, 11,
+     1, 0, 0, 1, 20,
+     0, 1, 1, 0, 22,
+     0, 1, 0, 1, 52,
+     0, 0, 1, 1, 55,
+     1, 1, 1, 0, 4,
+     1, 1, 0, 1, 13,
+     1, 0, 1, 1, 15,
+     0, 1, 1, 1, 38,
+     1, 1, 1, 1, 14))

---------------------------------------------------------------------
> openp(dat, dfreq=TRUE,m="up")
Model fit:
               deviance      df        AIC
fitted model      6.011       8     91.887

Test for trap effect:
                                    deviance      df        AIC
model with homogenous trap effect      3.555       7     91.431

## Note that capture probabilities are not constant.
Capture probabilities:
          estimate  stderr
period 1        --      --
period 2    0.4176  0.0517
period 3    0.4664  0.0334
period 4        --      --


## Not all Survival probabilities can be computed.
Survival probabilities:
               estimate  stderr
period 1 -> 2         1       0
period 2 -> 3         1       0
period 3 -> 4        --      --

Abundances:
          estimate  stderr
period 1        --      --
```

```
period 2      407.1     44.4
period 3      420.3     20.6
period 4         --        --

Number of new arrivals:
                  estimate   stderr
period 1 -> 2          --        --
period 2 -> 3        13.2      49.1
period 3 -> 4          --        --

Total number of units who ever inhabited the survey area:
              estimate   stderr
all periods      421.1        9

Total number of captured units: 385

-----------------------------------------------------------------
> openp(dat, dfreq=TRUE, m="ep")

Model fit:
                deviance      df       AIC
fitted model      43.867      11   123.743

Test for trap effect:
                                     deviance     df       AIC
model with homogenous trap effect       21.82     10   103.696

## Note that capture probabilities are constant.
Capture probabilities:
            estimate   stderr
period 1      0.5387    0.019
period 2      0.5387    0.019
period 3      0.5387    0.019
period 4      0.5387    0.019

## Survival probabilities can be computed.
Survival probabilities:
                  estimate   stderr
period 1 -> 2            1        0
period 2 -> 3            1        0
period 3 -> 4            1        0

Abundances:
            estimate   stderr
period 1       168.9     13.4
period 2       336.0     16.9
period 3       428.8      8.8
```

```
period 4      428.8      8.8

Number of new arrivals:
               estimate   stderr
period 1 -> 2      167.1     23.0
period 2 -> 3       92.8     19.4
period 3 -> 4        0.0      0.0

Total number of units who ever inhabited the survey area:
             estimate   stderr
all periods     428.8      8.8

Total number of captured units: 385
```

The data gives us a population estimate of 428.8 when the capture probabilities are constant, otherwise we get 421.1.