

Population Assessment and Distance Sampling

Zack Treisman

Spring 2021



Philosophy

Wildlife population assessment - *How many are there?*

Simpler questions:

- ▶ rank
- ▶ occupancy
- ▶ index

Population

- ▶ census
- ▶ mark-recapture
- ▶ removal
- ▶ plot sampling
- ▶ distance sampling

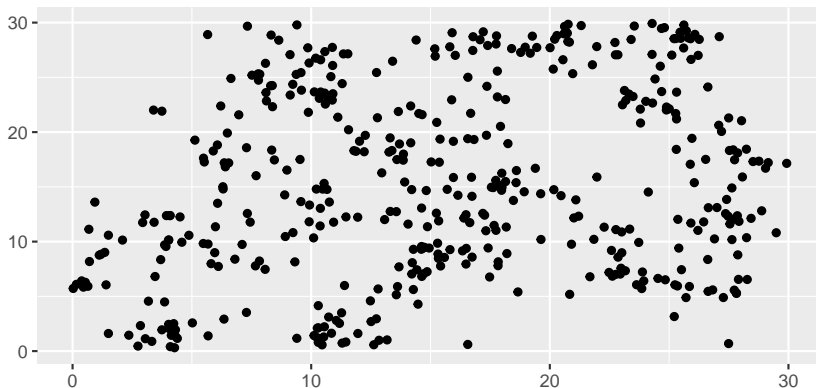
A simulated example

We'll use some of the code from the section on simulation to generate random arrangements of populations.

```
pos <- random_pop()
```

The simulated population size is 458.

```
ggplot(data=pos, aes(offspr_x, offspr_y))+geom_point()+labs(x="", y="")
```



Capture-mark-recapture

- ▶ Capture a subset of the population and mark them.
- ▶ Release. After sufficient time for the released subset to have fully mixed with the population, capture another subset.
- ▶ The fraction of the second capture that is marked estimates the fraction of the population in the first capture.
- ▶ As a formula: $\hat{P} = C_1/p_m$ where
 - ▶ \hat{P} is the population estimate,
 - ▶ C_1 is the size of the first captured group and
 - ▶ p_m is the marked proportion in the second captured group.

```
capture1 <- sample(nrow(pos), 50)
pos$mark <- 1:nrow(pos) %in% capture1

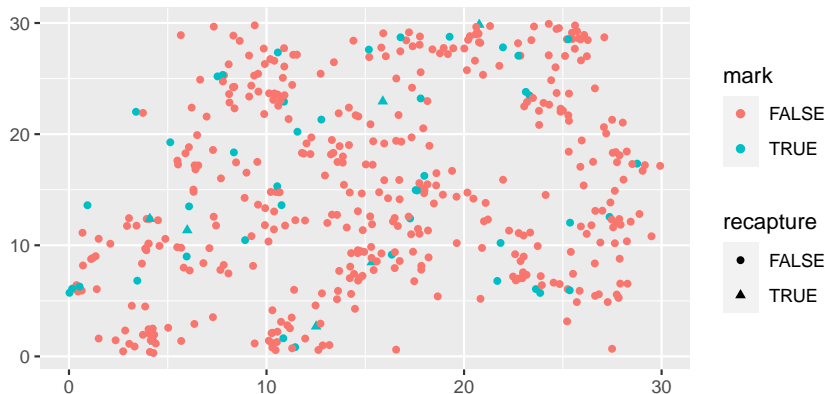
capture2 <- sample(nrow(pos), 50)
pos$recapture <- (1:nrow(pos) %in% capture2) & pos$mark
```

- ▶ There are 7 recaptures.
- ▶ This is 0.14 of the second captured group.
- ▶ The population estimate is $50/0.14=357$.

Plot of captures and recaptures

This method is not particularly well suited to populations that don't move around and mingle.

```
ggplot(data=pos, aes(offspr_x, offspr_y, color=mark, shape=recapture))+  
  geom_point()+labs(x="", y="")
```



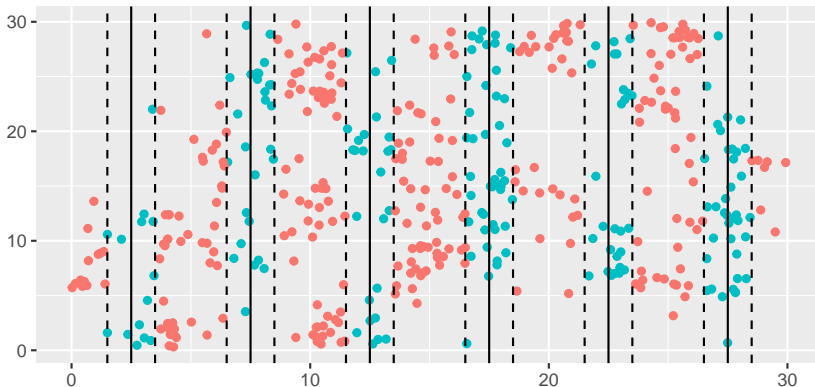
Capture-mark-recapture is fundamentally Bayesian

Population estimates can be improved over time by repeating the experiment.

Plot Sampling

Even if the individuals being counted don't move around, it is often impractical/unreasonable/impossible to census an entire area of interest. Instead, we often census appropriately chosen subplots.

- Subplots are often laid out along transects.



Computing population from plot sampling

Divide the number counted in the plots by the fraction of the area that is covered by the plots.

```
num_plots <- 5
half_width <- 1
Lx <- 30; Ly <- 30
counted <- sum(pos$in_plot)
area_plots <- Ly*2*half_width*num_plots
total_area <- Lx*Ly
area_fraction <- area_plots/total_area
(pop_estimate <- counted/area_fraction)

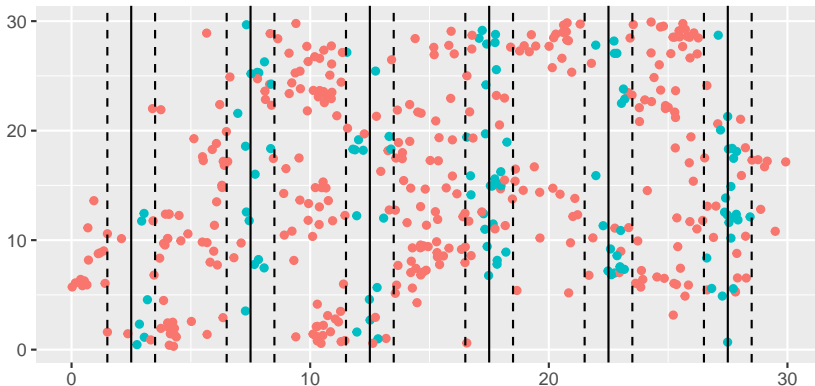
## [1] 528
```

We can use a bootstrap to get a confidence interval for this estimate.

Distance sampling

What if detecting the individuals in the population is challenging?

- In many surveys, the further the individual is from the observer, the lower the probability of observation.



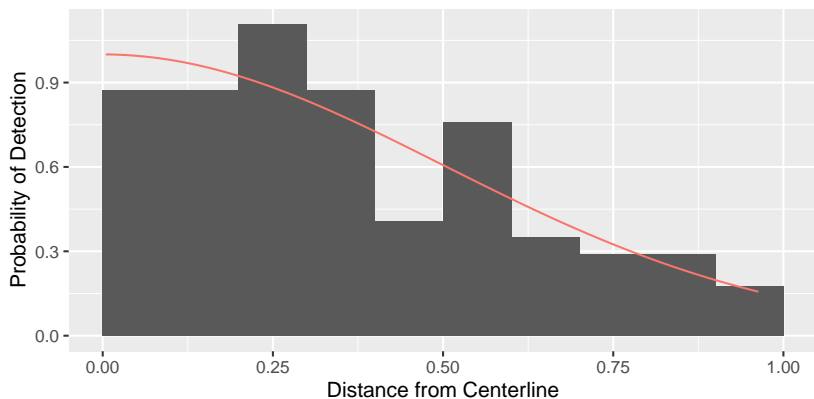
```
counted <- sum(pos$detected)
(pop_estimate <- counted/area_fraction)
```

```
## [1] 309
```

Overall detection probability

It is useful to look at a histogram of the distance of detected individuals.

- Goal: fit a detection probability function to these data.



The overall probability of detection is $\hat{P} = 0.5981$.

Estimation using \hat{P}

To get an estimated count from a count where the probability of detection is not 100%, divide the actual count by the overall probability of detection.

```
est_count <- counted/overall_prob  
(pop_estimate <- est_count/area_fraction)
```

```
## [1] 516.598
```

Guessing the detection probability function

We have some data that we would like to use to define a curve.

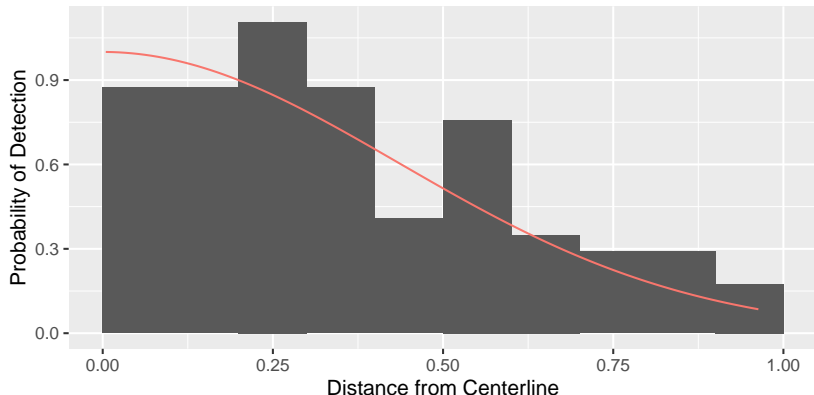
- ▶ Propose a shape for the detection probability function (half-normal, hazard rate, uniform, ...).
 - ▶ $p_{det}(0) = 1$ (objects at the centerline are detected)
 - ▶ $p'_{det}(0) = 0$ (detection probability is rather flat around 0)
 - ▶ $p'_{det}(x) \leq 0$ (detection probability decreases with distance)
- ▶ Use maximum likelihood to find parameters of the most appropriate curve in the family.

The \hat{P} above comes from rescaling a normal distribution with $\sigma = 0.5$ so that $p_{det}(0) = 1$. The choice of σ was from fitting by eye. The code below finds the σ for the maximum likelihood.

```
samp_ind<-pos[pos$in_plot,]  
prob_detect2 <- function(x,s) dnorm(x,sd=s)/dnorm(0, sd=s)  
nll_fn<-function(s) -sum(dbinom(samp_ind$detected,  
                                1,prob_detect2(samp_ind$d,s),  
                                log = TRUE))  
opt1 <- optim(fn = nll_fn, par = list(s=0.5), method="BFGS"); opt1$par
```

```
##           s  
## 0.4852794
```

Using the optimal half-normal detection probability



```
overall_prob2 <- integrate(function(x) prob_detect2(x,s=opt1$par),0,1)$value  
est_count2 <- counted/overall_prob2  
(pop_estimate2 <- est_count2/area_fraction)
```

```
## [1] 528.8525
```

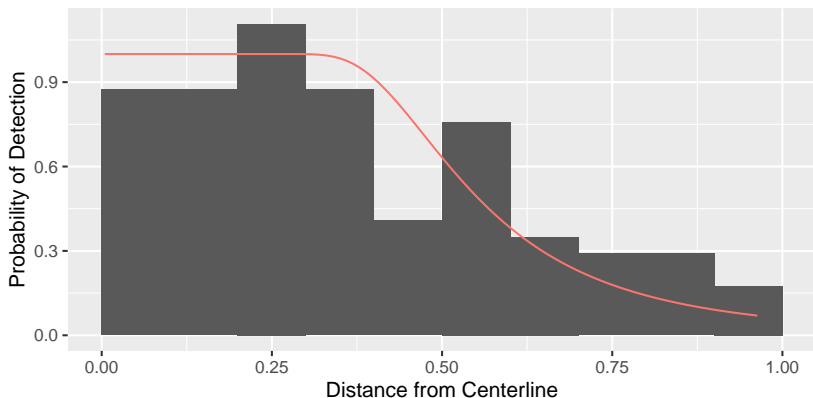
Hazard rate detection probability functions.

A half-normal detection probability function may not be best.

The hazard rate detection probability is defined by

$$p_{det}(x) = 1 - \exp(-(x/\sigma)^{-b}).$$

This function has two parameters, σ which controls the width of the shoulder, and b which controls the steepness of the dropoff.



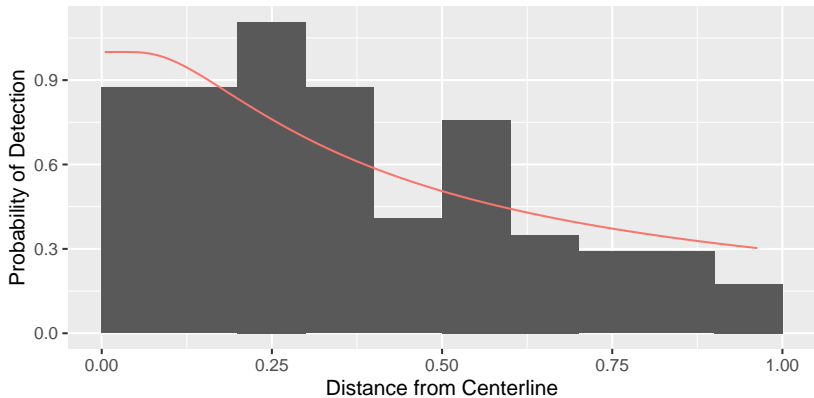
Maximum likelihood to find the optimal hazard rate

Again, we can use maximum likelihood to estimate the parameters.

```
prob_detect_hr <- function(x,s,b) 1-exp(-(x/s)^(-b))
nll_fn_hr <- function(pars){s<-pars[1];b<-pars[2]
-sum(dbinom(samp_ind$detected,1,prob_detect_hr(samp_ind$d,s,b), log = TRUE))}
opt2 <- optim(fn = nll_fn_hr, par = list(s=0.25, b=4)); opt2$par
```

```
##           s           b
## 0.3546658 1.0210627
```

The optimal Hazard Rate detection curve



Using the optimal Hazard Rate function to predict \hat{P}

And then we estimate

- ▶ the overall probability \hat{P} by integrating the detection probability function,
- ▶ the number of individuals within the sampling region by dividing the count by \hat{P} , and
- ▶ the population by dividing the number of individuals by the area fraction.

```
overall_prob3<-integrate(function(x){  
  prob_detect_hr(x,s=opt2$par[1],b=opt2$par[2])},0,1)$value  
est_count3 <- counted/overall_prob3  
(pop_estimate3 <- est_count3/area_fraction)
```

```
## [1] 537.1223
```

Assumptions and issues

- ▶ If the individuals that are being counted are clustered (pods of whales, clusters of plants, . . .), this can cause problems.
 - ▶ Estimate the distance to the center of the cluster.
 - ▶ Record the size of each cluster and estimate the mean cluster size.
- ▶ Often, instead of using all observations, it makes sense to truncate observations above some effective strip width.
- ▶ If there is avoidance of the observer, this can cause problems.

Point-transect distance sampling

For studies where the observer is stationary, the process is quite similar.

References