

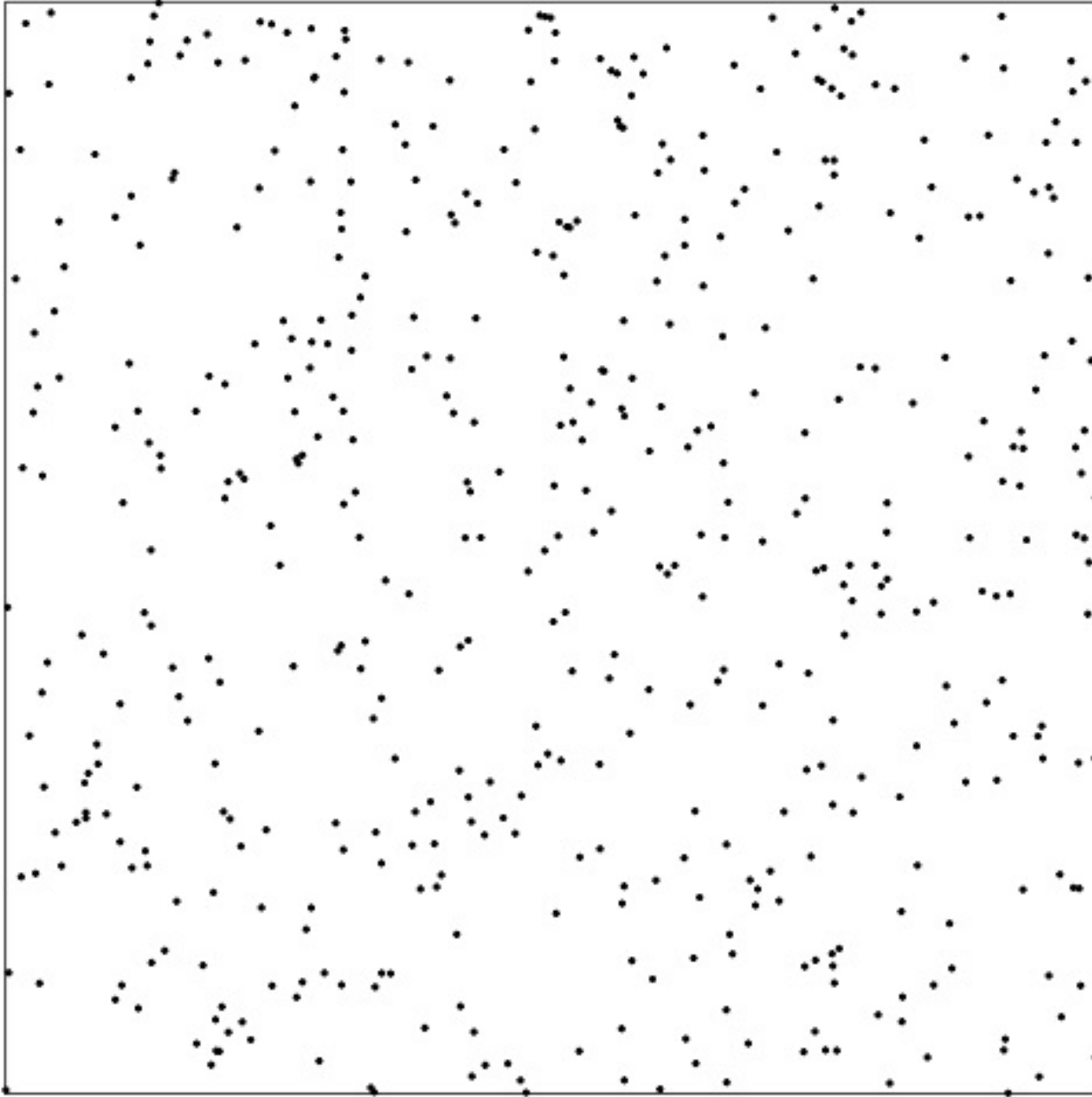
# Introduction to distance sampling

David L Miller

# Overview

- Line transects
- Simple estimates of abundance
- Why is detectability important?
- What is a detection function?
- First look at fitting models in R

# How many animals are there? (500!)



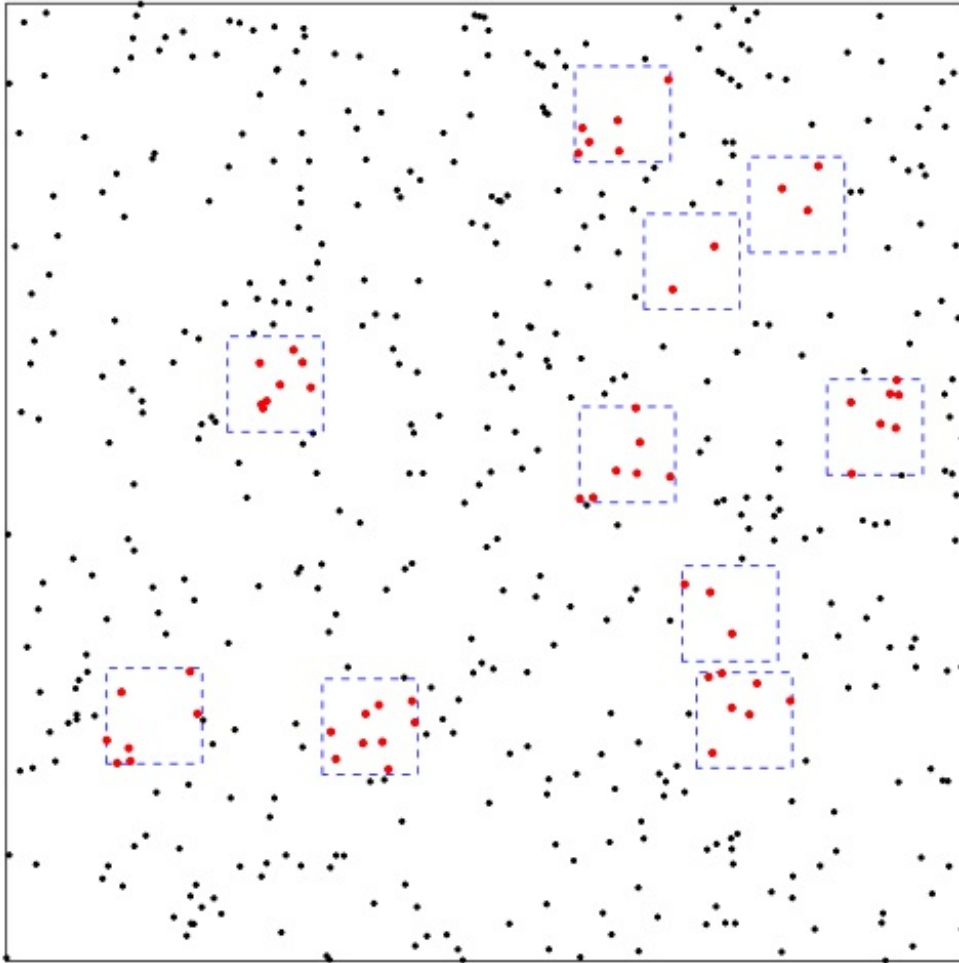
# General strategy

- Take a sample in some fixed areas
- Find density/abundance in *covered area*
- Multiply up to get abundance

# General strategy (What did we assume?)

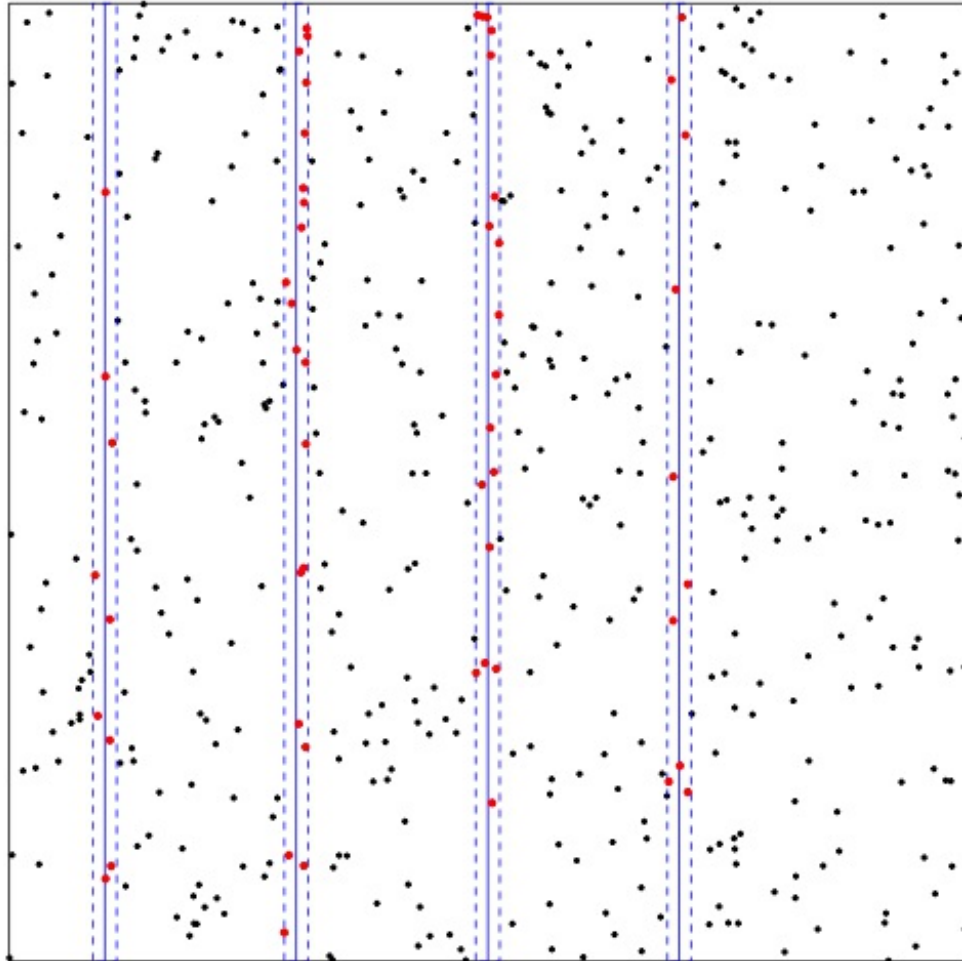
- Take a sample in some fixed areas
  - *Sample is representative*
- Find density/abundance in *covered area*
  - *Estimator is “good”*
- Multiply up to get abundance
  - *Sample is representative*

# Plot sampling



- Surveyed 10 quadrats (each  $0.1^2$  units)
  - Total covered area  
 $a = 10 * 0.1^2 = 0.1$
- Saw  $n = 59$  animals
- Estimated density  
 $\hat{D} = n/a = 590$
- Total area  $A = 1$
- Estimated abundance  $\hat{N} = 590$

# Strip transect



- Surveyed 4 lines (each  $1 * 0.025$  units)
  - Total covered area  
 $a = 4 * 1 * 0.025 = 0.1$
- Saw  $n = 57$  animals
- Estimated density  
 $\hat{D} = n/a = 570$
- Total area  $A = 1$
- Estimated abundance  $\hat{N} = 570$

# Detectability



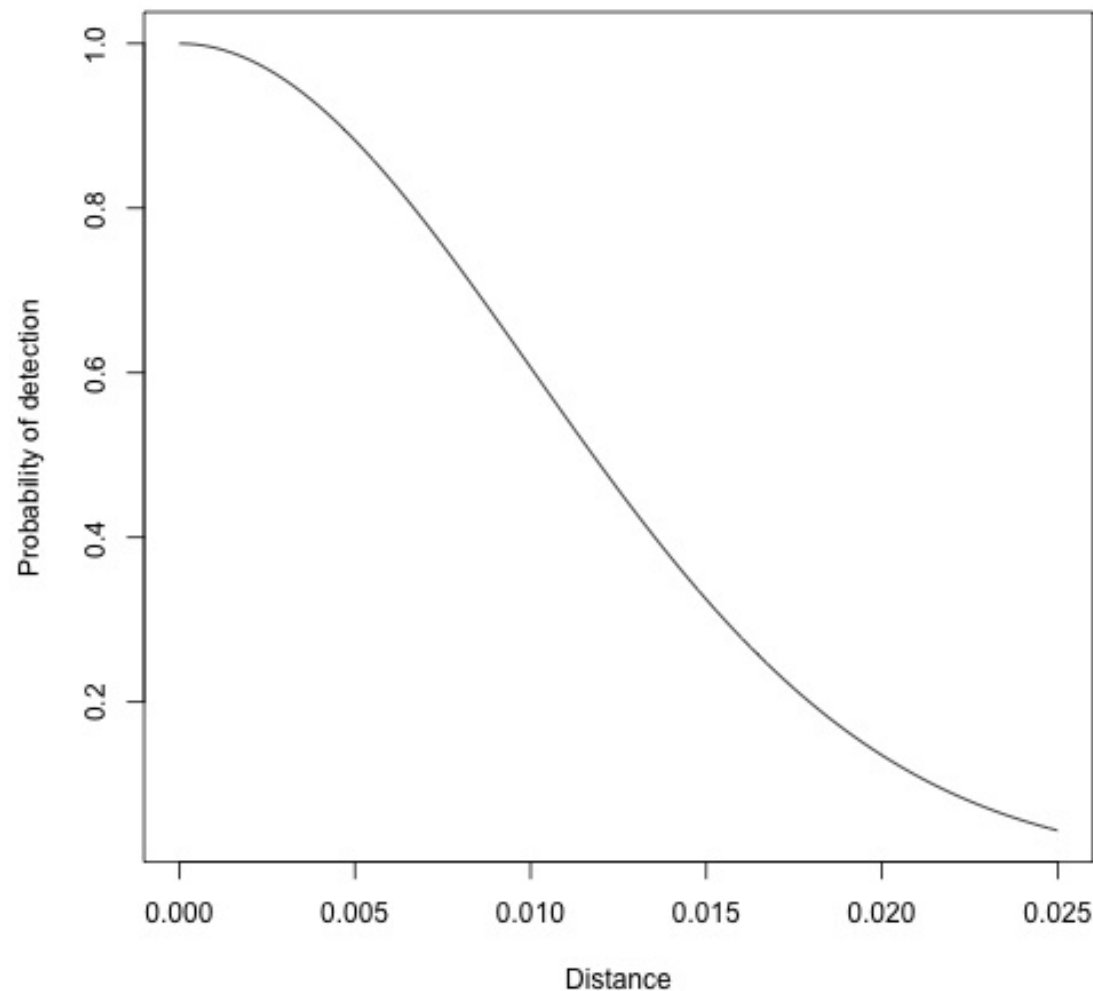
# Detectability matters!

- We've assumed certain detection so far
- This rarely happens IRL
- Distance to the **line** is important
  - (Other things too, more on that later)
  - Detectability should decrease with increasing distance

# Recording distances is more efficient

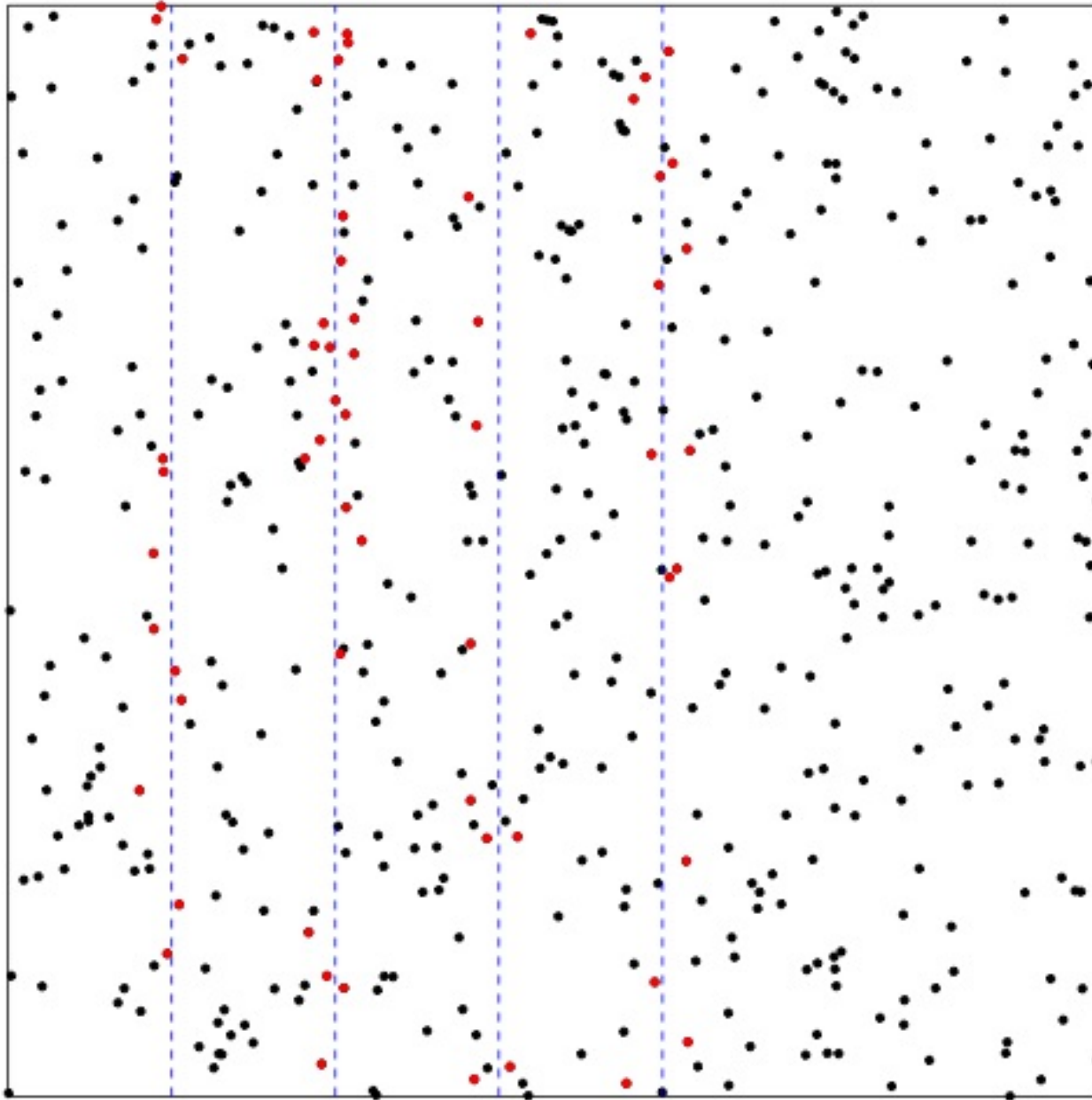
- Plots: what if an animal is *just* outside the box?
- Strips: what if an animal is *just* outside the strip?
- Line transects: record **everything**, then discard later
  - Decide strip width (*truncation distance*) later

# Detection as a function of distance

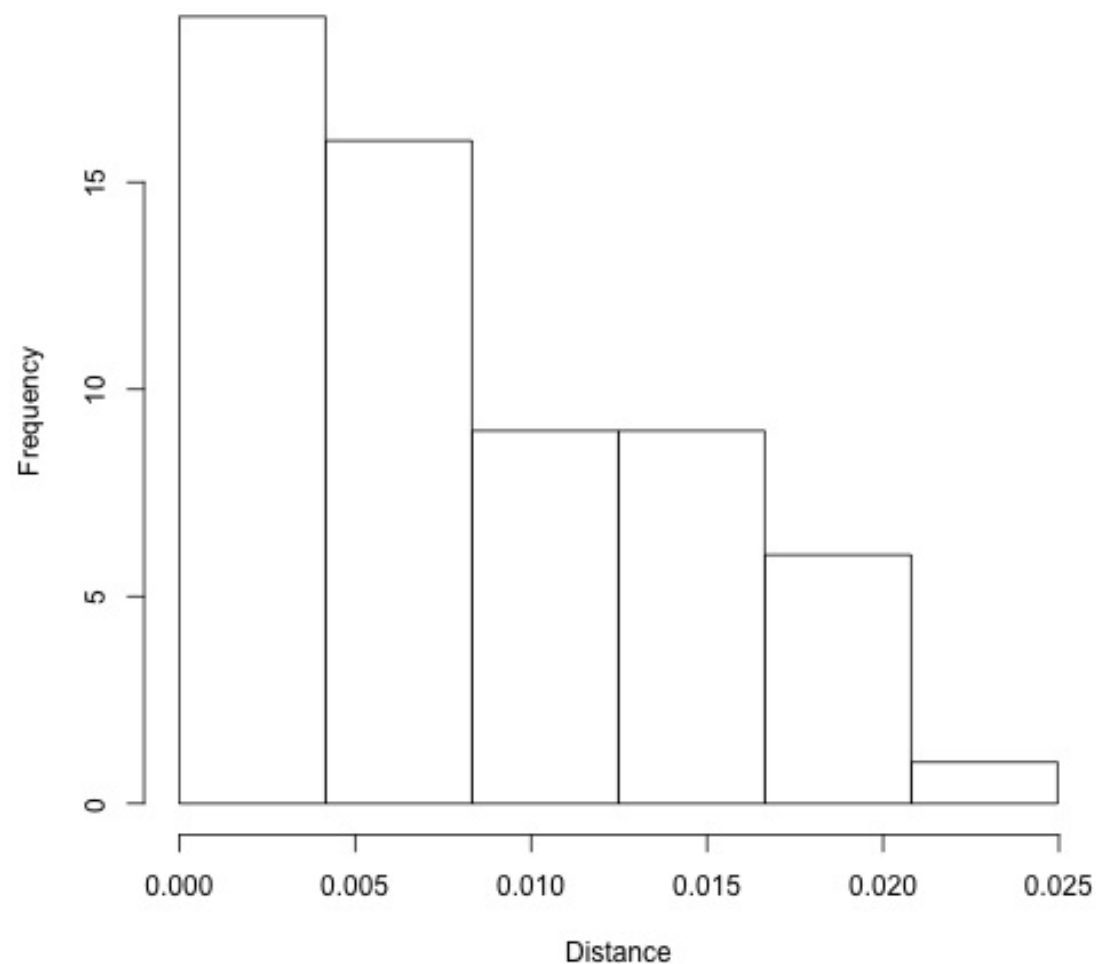


- Model probability of detection, given distance
- Fit models for the curve
- Derive a probability of detection from this model

# Line transect



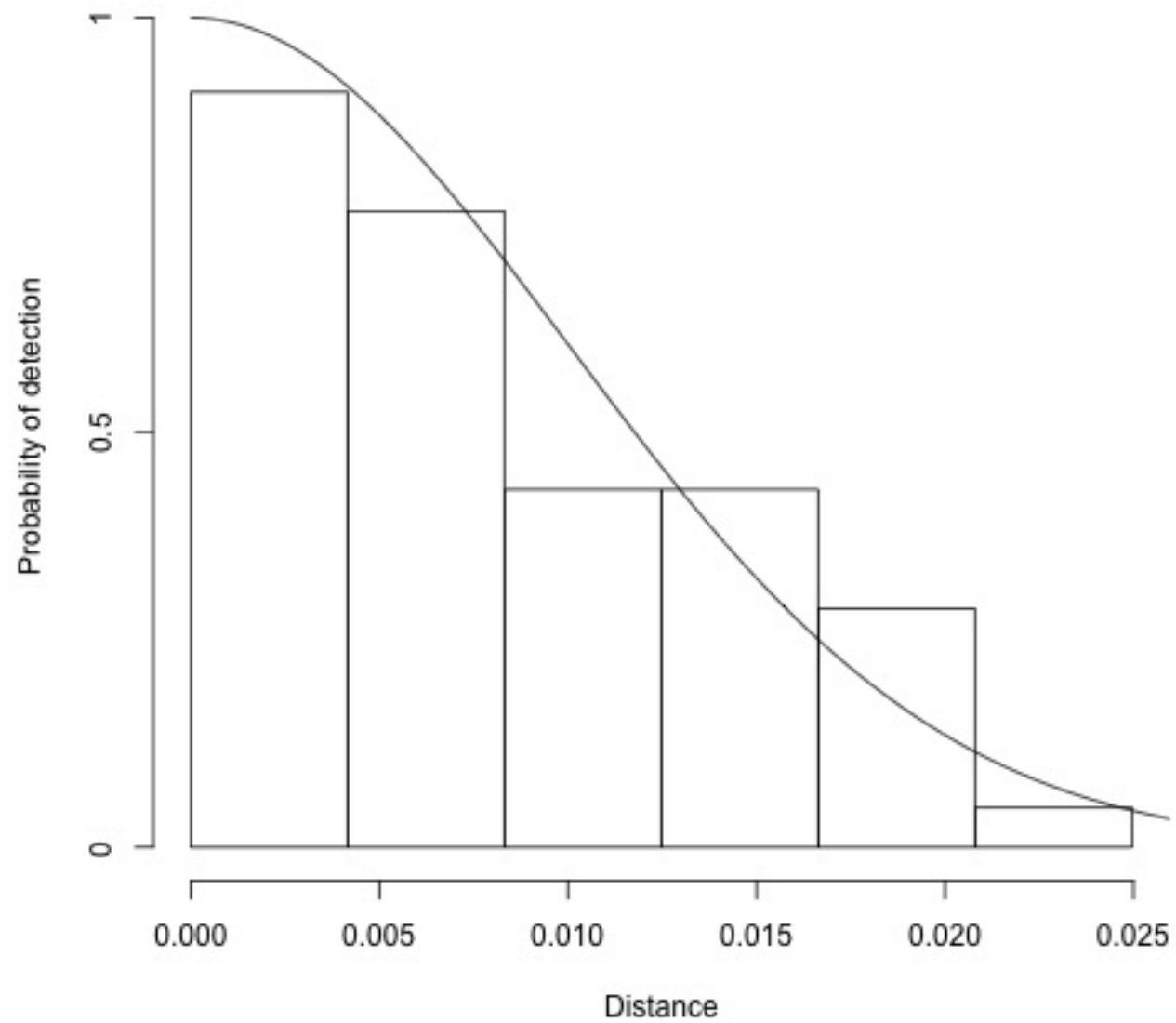
# Line transects - distances



- Now we recorded distances, what do they look like?
- “Fold” distribution over, left/right doesn't matter
- Drop-off in # observations w. increasing distance

“You should model that”

# Detection function



# Using distance information

- Detection function:  $\mathbb{P}(\text{detection} \mid \text{at distance } x)$
- Integrate out the conditioning  $\Rightarrow \mathbb{P}(\text{detection}) = \hat{p}$
- “Inflate”  $n$  by  $\hat{p}$  to estimate abundance



# Distance sampling estimate

- Surveyed 5 lines (each  $1 * 0.025$  units)
  - Total covered area  $a = 5 * 1 * 0.02 = 0.2$
- Probability of detection  $\hat{p} = \int_0^w \frac{g(x)}{w} dx = 0.5981$
- Saw  $n = 60$  animals
- Inflate to  $n/\hat{p}$
- Estimated density  $\hat{D} = \frac{n/\hat{p}}{a} = 502$
- Total area  $A = 1$
- Estimated abundance  $\hat{N} = 502$

# Summary: line transects

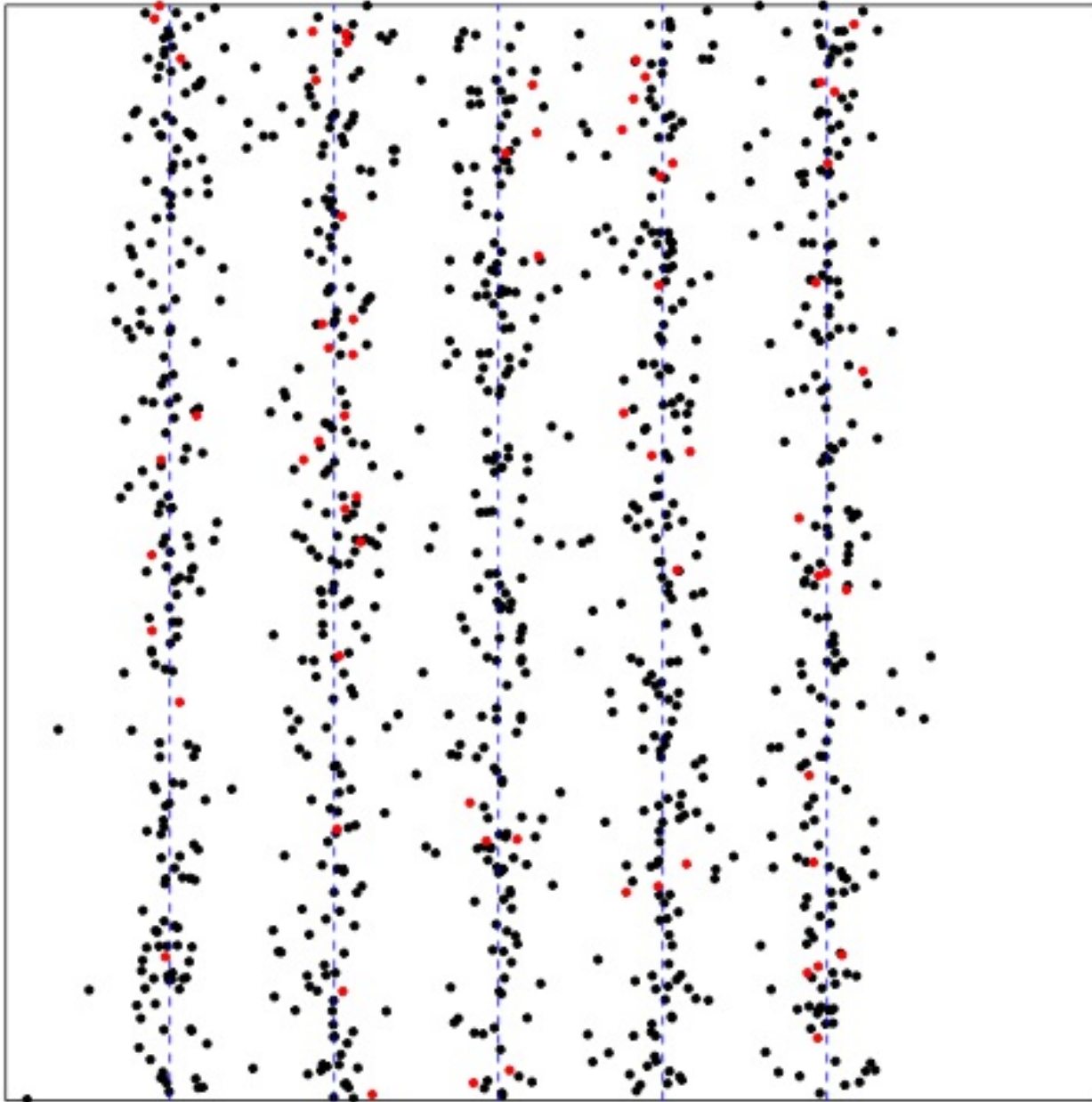
- Efficient survey design
- Relax the assumption of perfect detection
- Exchange assumptions for data
- More information = better inference

# Assumptions

# Assumptions

1. Animals are distributed independent of lines
2. On the line, detection is certain
3. Distances are recorded correctly
4. Animals don't move before detection

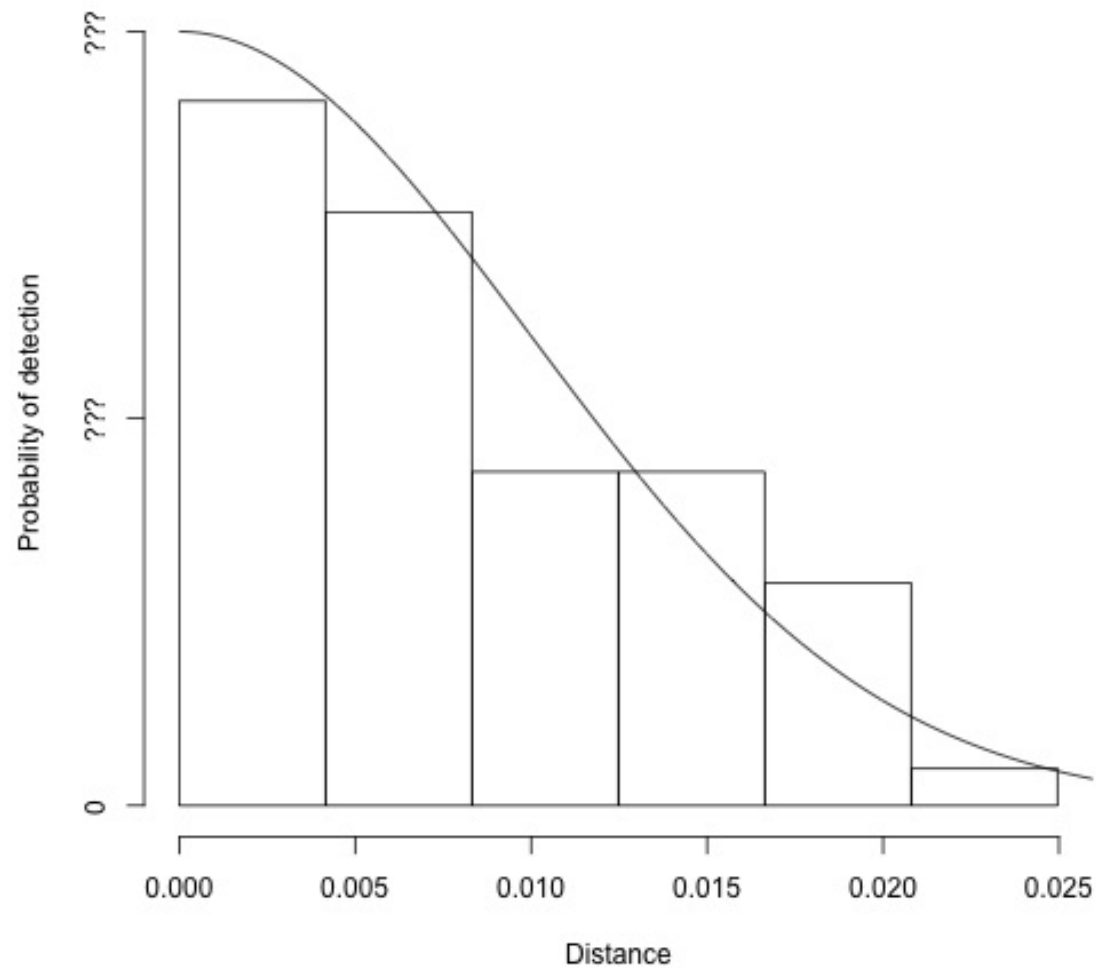
# Animals are distributed independent of lines



- When transects follow features
- Difficult to work out detectability vs. distribution

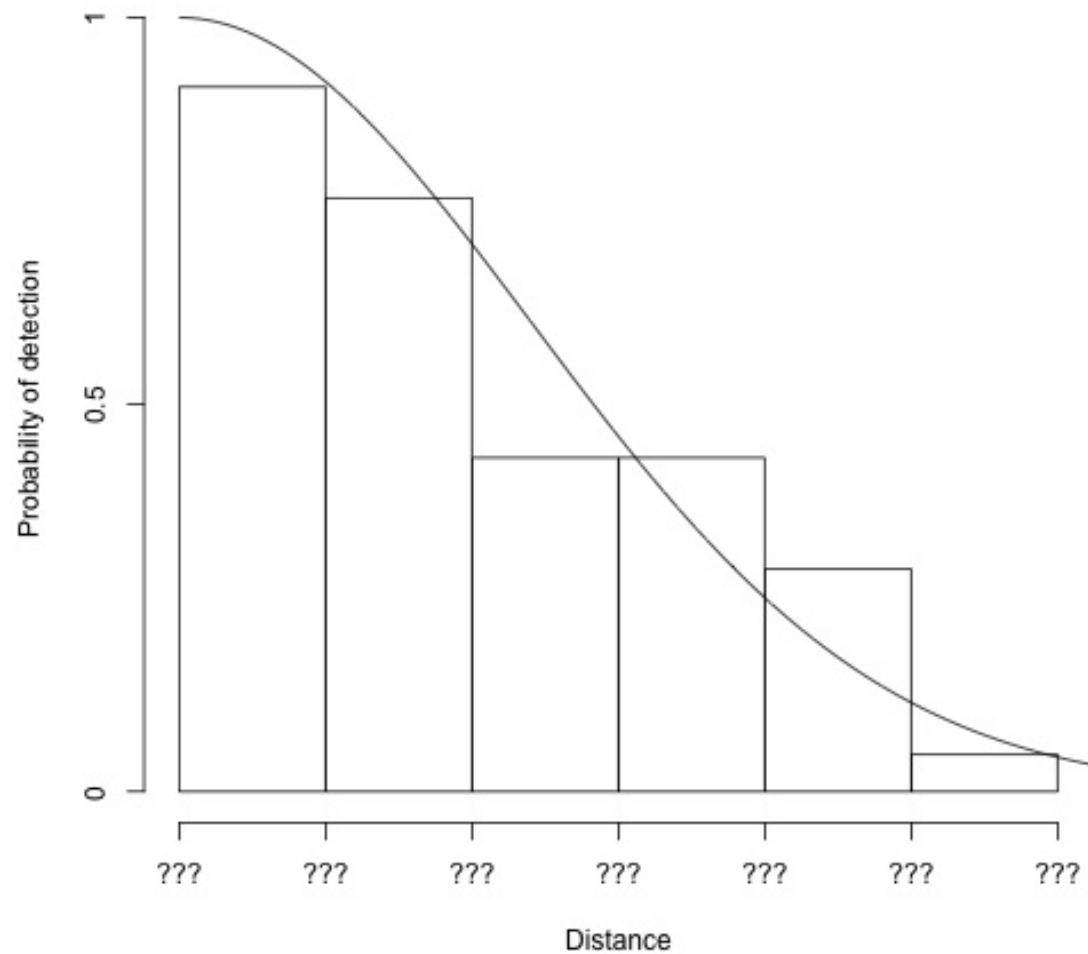
# On the line, detection is certain

- Perception bias
- Availability bias
- Don't know y axis scale

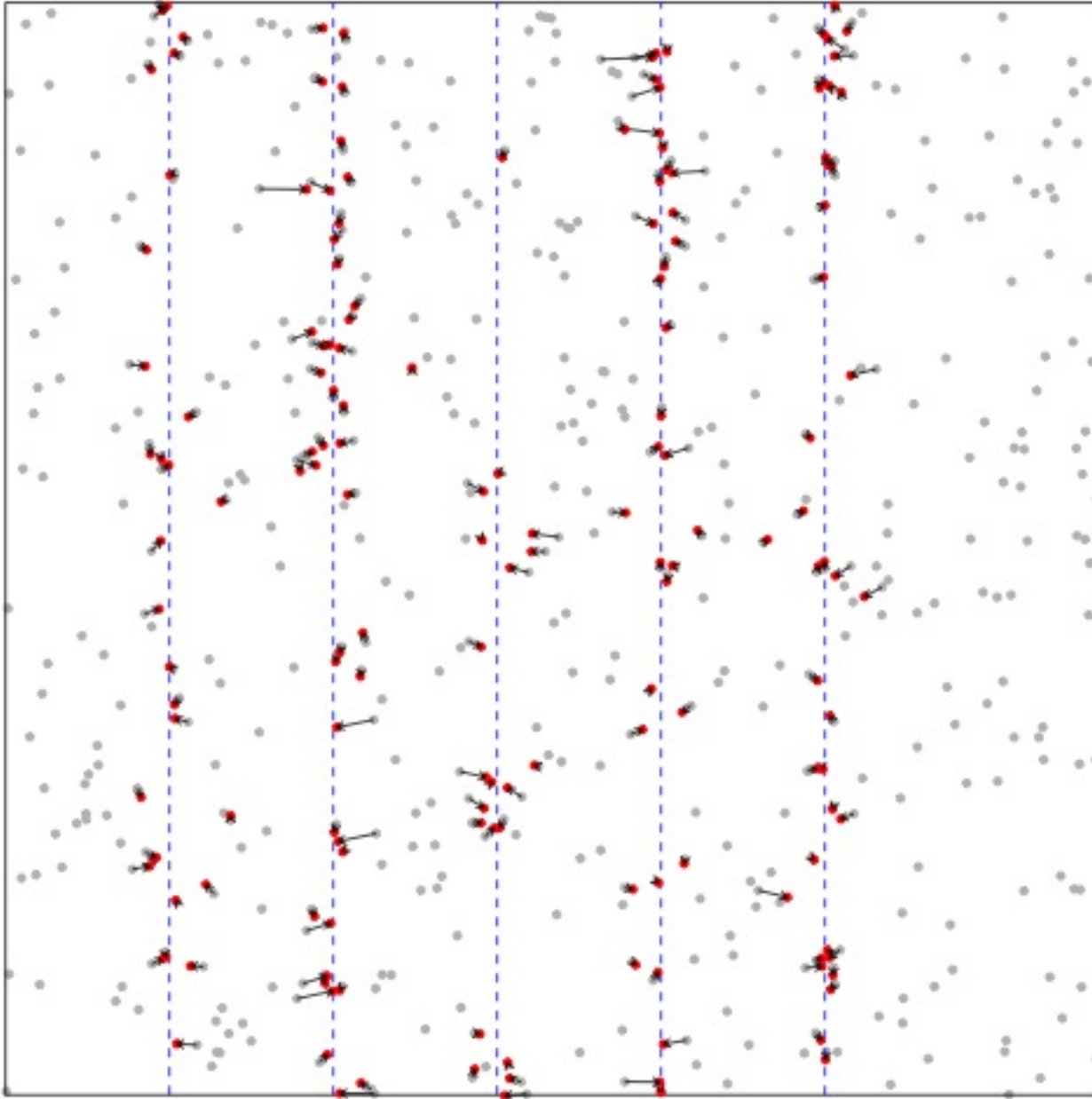


# Distances are recorded correctly

- Measurement error
- Don't know X axis scale
- This can be systematic



# Animals don't move before detection



- Animals can be attracted or repelled
- Problems with distribution wrt line and/or measurement error



# Detection functions

# What are detection functions?

- Model  $\mathbb{P}$  (detection | animal at distance  $x$ )
- (Hence the integration)
- Many different forms, depending on the data
- All share some characteristics

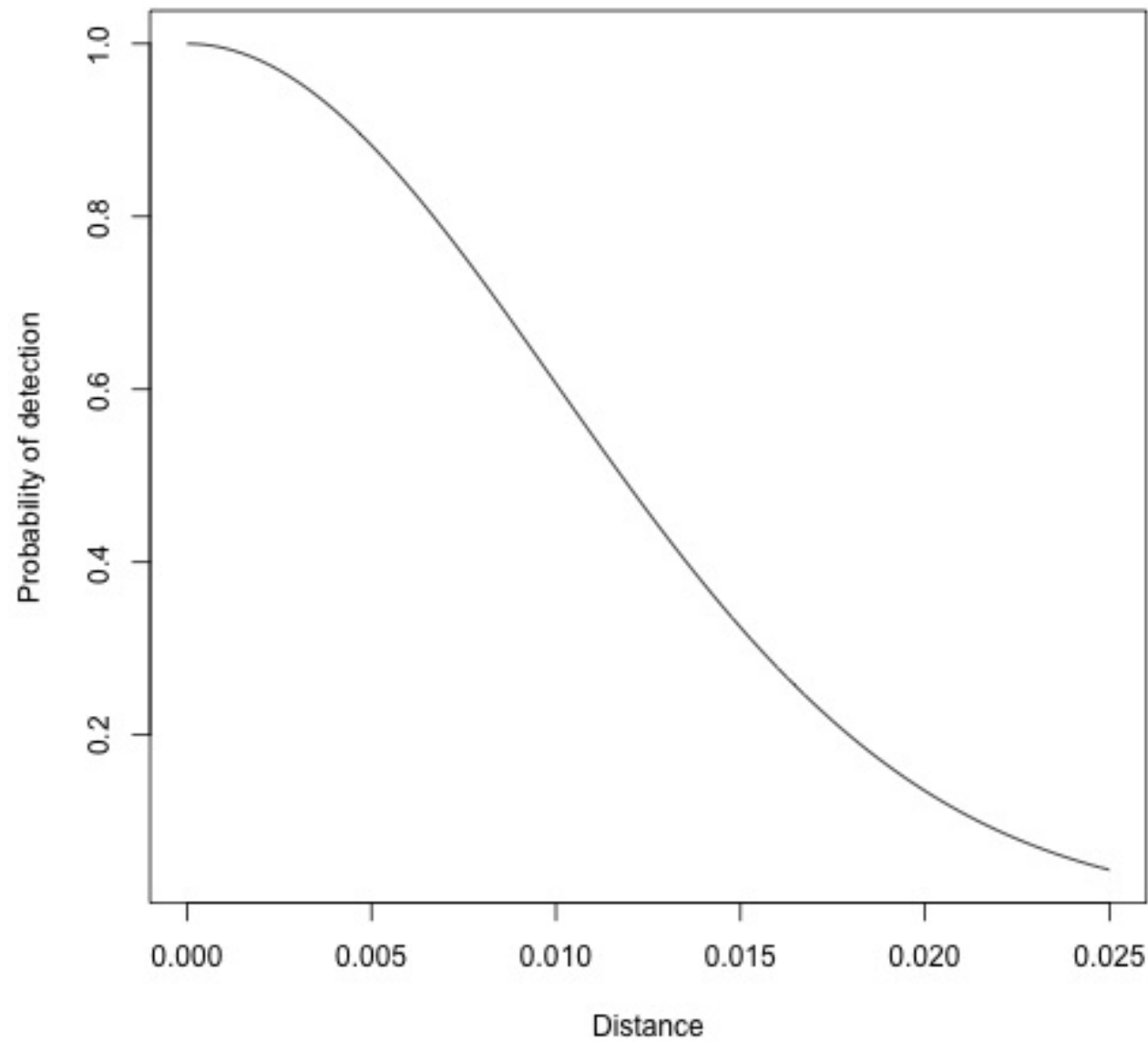
# Detection function assumptions

- Have a “shoulder”
  - *we see things nearby easily*
- Monotonic decreasing
  - *never increasing with increasing distance*
- “Model robust”
  - *lots of forms/flexible models*
- “Pooling robust”
  - *individual heterogeneity averages out*
- “Efficient”
  - *models don't need lots of parameters*

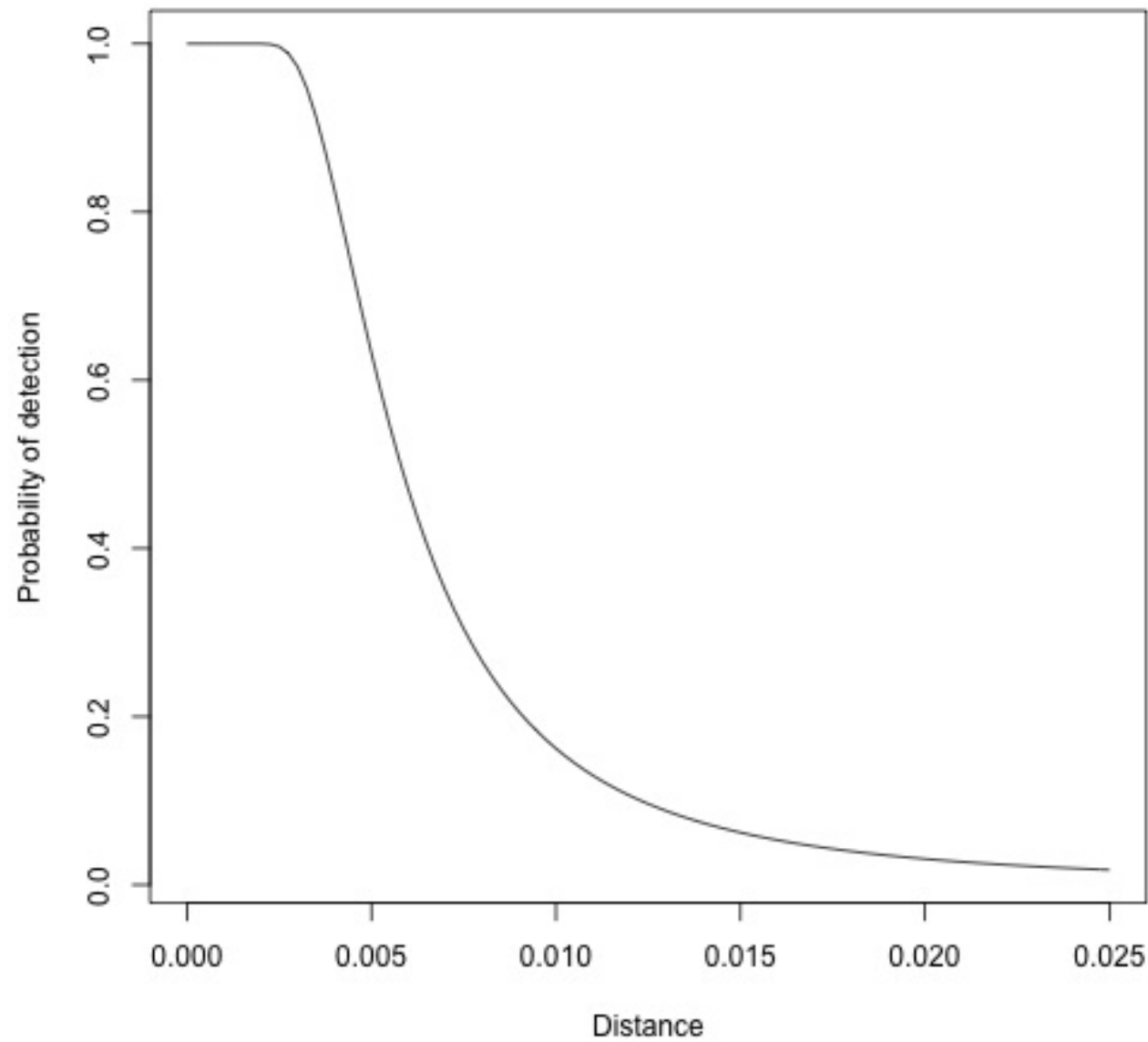
# Possible detection functions

- There are many options
- A restricted set we'll cover in this course...
  - Half-normal
  - Hazard-rate
  - adjustments to the above

# Half-normal detection functions



# Hazard-rate detection functions



# Adjustment terms

- These models are flexible
- What about adding more flexibility by “adjusting” them
- Options:
  - Cosine series
  - Polynomials
  - Hermite polynomials
- Add extra flexibility

Okay, but how can we actually  
do this?



# Modelling strategy

1. Pick some formulations, fit models
2. Check assumptions are violated
3. Goodness of fit
4. Select models
5. Estimate  $\hat{N}$  (and uncertainty!)

# Distance sampling data

- Need to have data setup a certain way
  - a `data.frame` with one row per observation
  - at least 2 columns, named “object” and “distance”

	distance	object	size	SeaState
1	246.0173	1	2	3.0
2	1632.3934	2	2	2.5
3	2368.9941	3	1	3.0
4	244.6977	4	1	3.5
5	2081.3468	5	1	4.0
6	1149.2632	6	1	2.4

# Fitting detection functions (in R!)

- Using the package `Distance`
- Function `ds()` does most of the work

```
library(Distance)
df_hn <- ds(distdata, truncation=6000, adjustment = NULL)
df_hr <- ds(distdata, truncation=6000, key="hr", adjustment =
NULL)
```

# Model summary

```
summary(df_hn)
```

Summary for distance analysis

Number of observations : 132

Distance range : 0 - 6000

Model : Half-normal key function

AIC : 2252.06

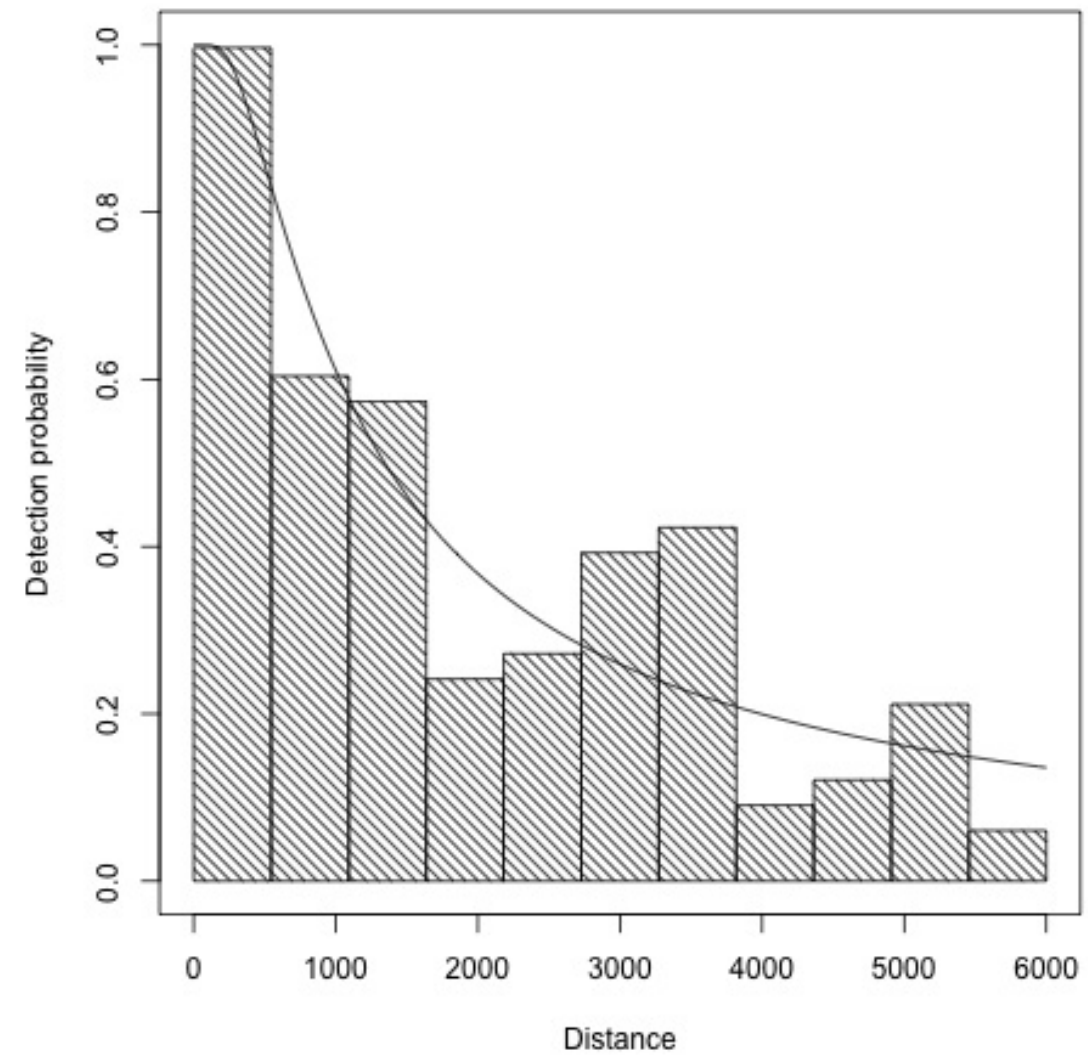
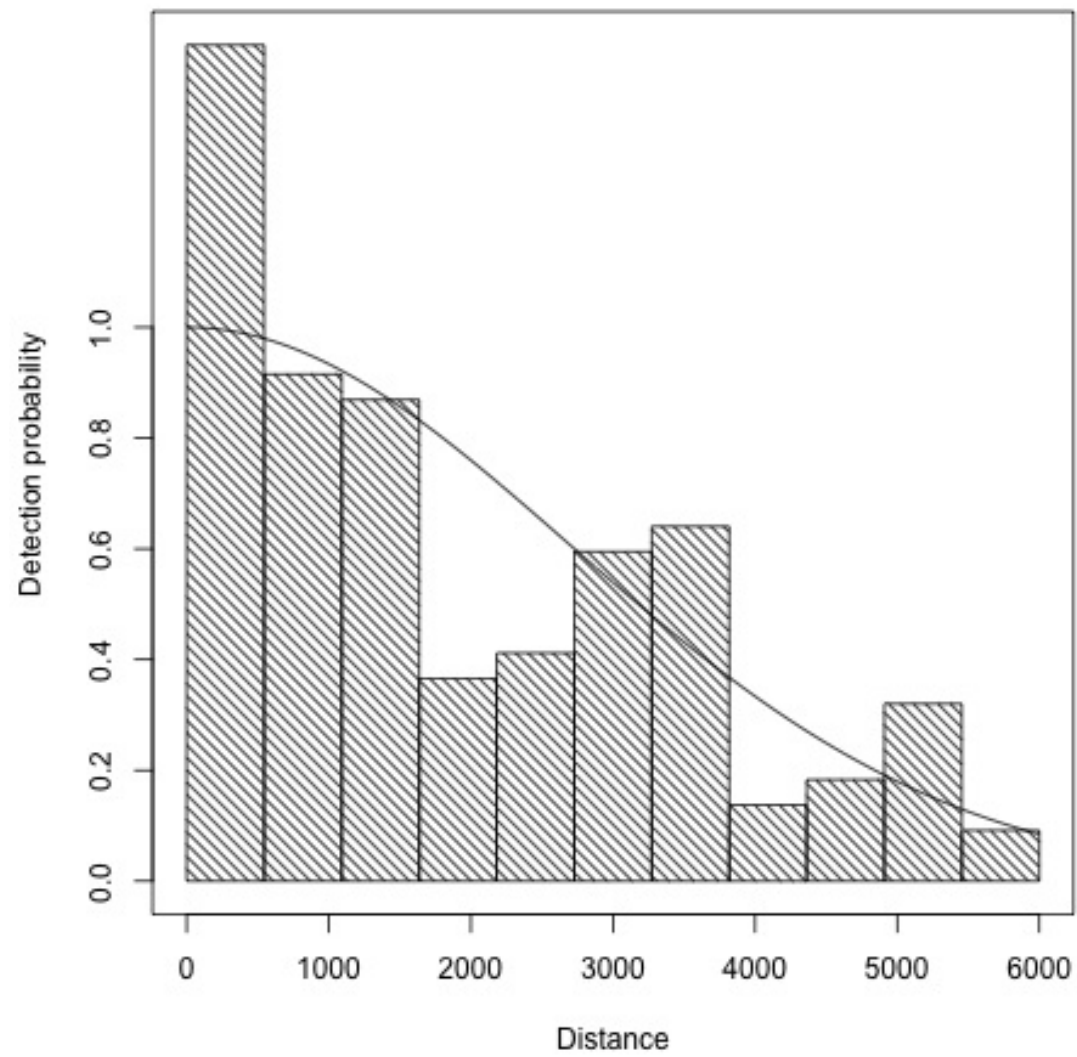
Detection function parameters

Scale Coefficients:

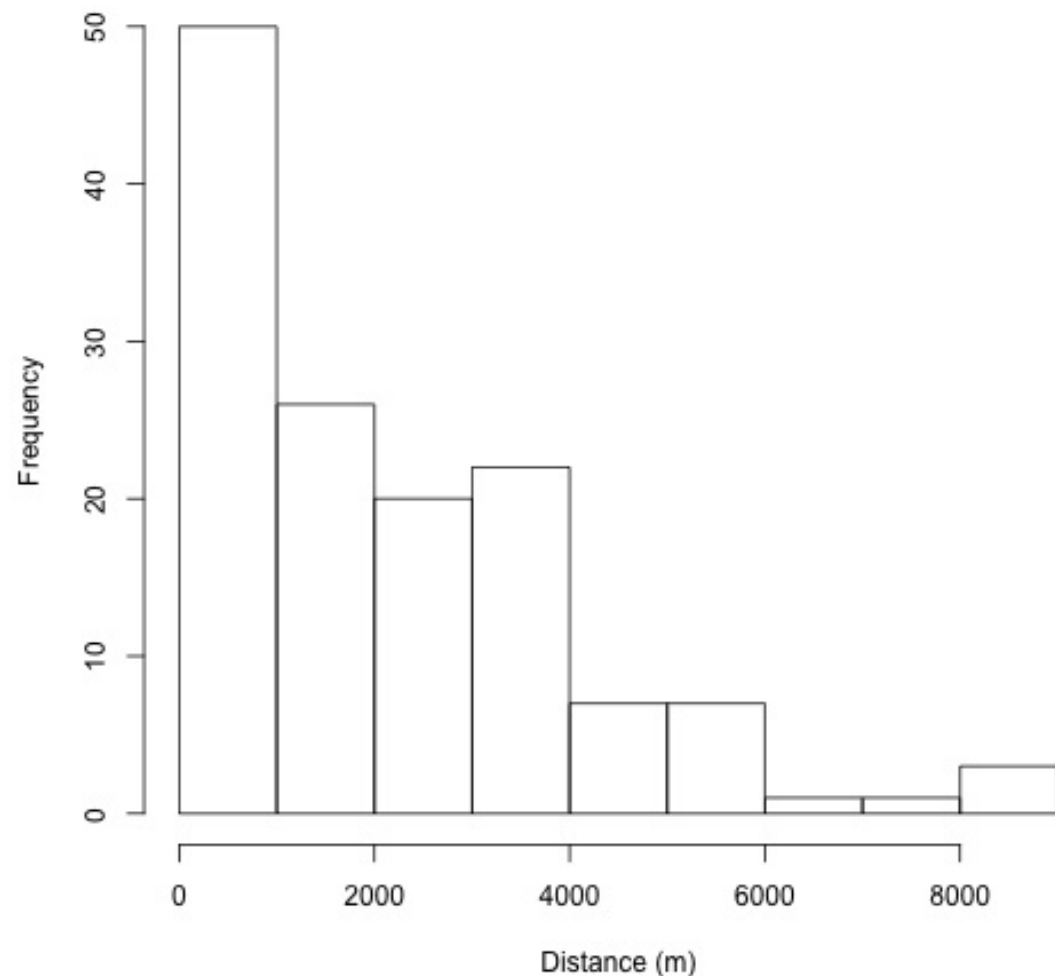
	estimate	se
(Intercept)	7.900732	0.07884776

	Estimate	SE	CV
Average p	0.5490484	0.03662569	0.06670757
N in covered region	240.4159539	21.32287580	0.08869160

# Plotting models



# Truncation



- We set `truncation=6000`, why?
- Remove observations in the tail of the distribution
- **Care about g near 0!**
- Trade-off! (Here we use ~96% of the data)
- Len Thomas suggests  $g(w) \approx 0.15$

# Recap

# Distance sampling

- More efficient sampling
  - No census
- Collect additional information
  - Distances
- Estimate detection
- Use  $\mathbb{P}(\text{detection})$  to correct counts



# What's next?

- Model checking and selection
- Estimating abundance in R
- Stratification
- What else affects detectability?