Introduction to Mark-Recapture Distance Sampling (MRDS)

- The "g(0) problem": missing animals on the transect line
- Intuitive introduction to mark-recapture distance sampling (MRDS)
- Full independence and point independence models
- Double observer configurations
- Assumptions and conclusions

For more information, see:

- Laake et al. (2004) – chapter in Advanced Distance Sampling book first describing the methods
- Burt et al. (2014) – accessible introduction to MRDS

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Conventional distance sampling

E.g., line transects

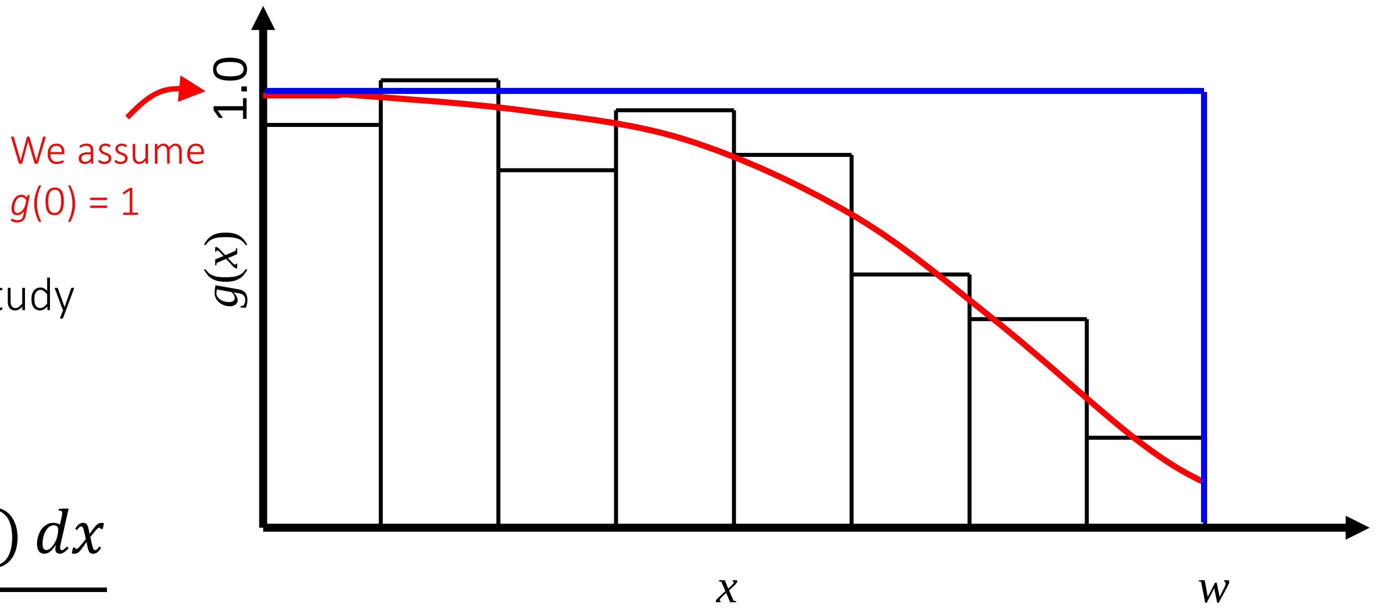
animals detected

$$\hat{N} = \frac{n}{\hat{P}_a} \times \frac{A}{2wL}$$

detection probability

proportion of study area surveyed

$$\hat{P}_a = \frac{\text{area under curve}}{\text{area under rectangle}} = \frac{\int_0^w \hat{g}(x)\,dx}{1 \times w}$$



Fundamental assumption: every animal on the transect line is detected – i.e., $g(0) = 1$

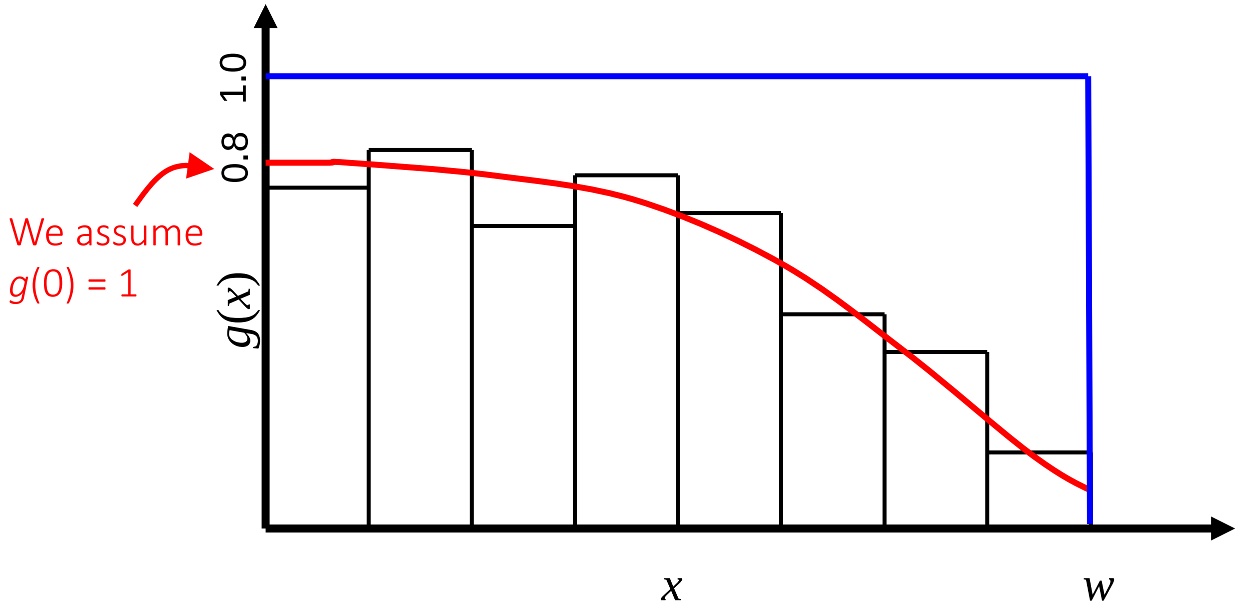
What if g(0)<1?

If g(0)<1 we get a negative bias in estimates of N (and D)

E.g., if g(0)=0.8 then estimates of N and D are 80% of true value on average

Nothing in the perpendicular distance data to tell us g(0)<1

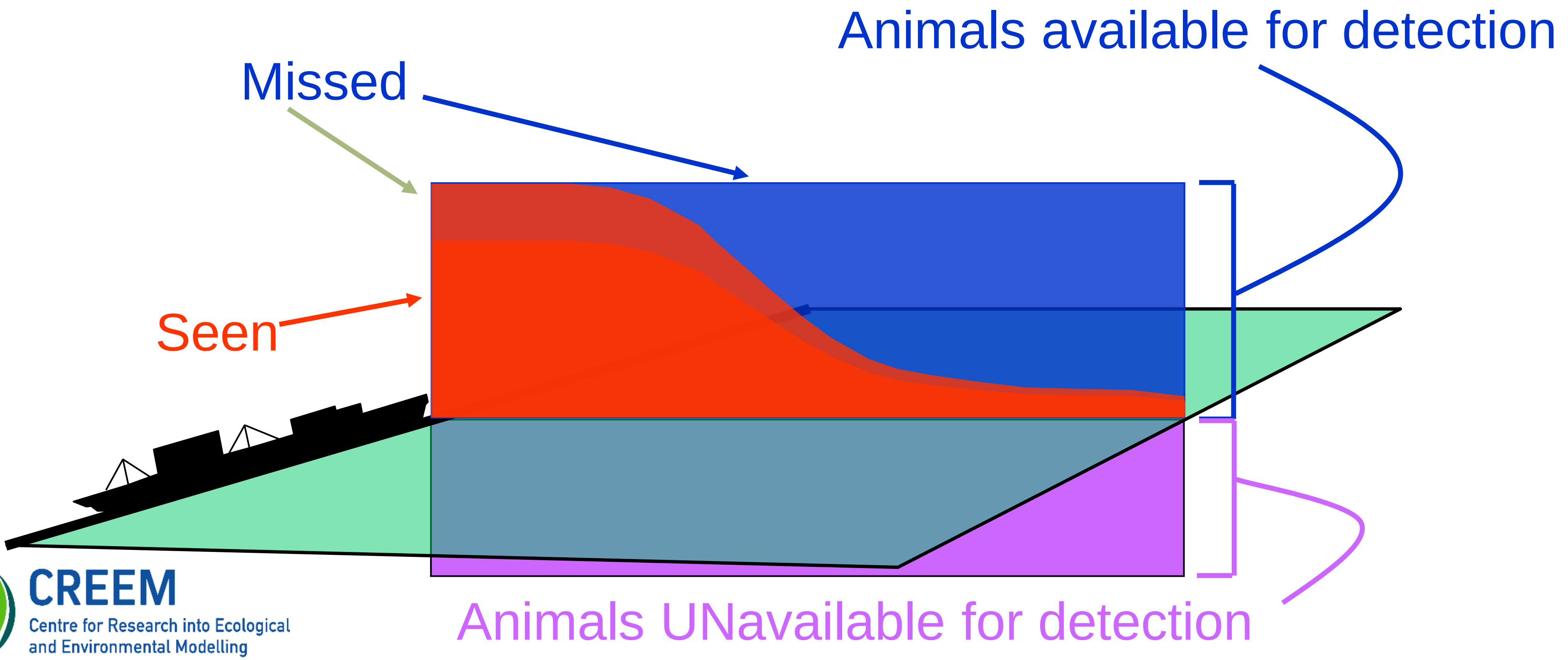
Additional data are needed. This talk is about one approach for what data to collect and how to analyse it.



$$\hat{P}_a = \frac{\int_0^w \hat{g}(x)\,dx}{1 \times w} \qquad \hat{N} = \frac{n}{\hat{P}_a} \times \frac{A}{2wL}$$

Availability and perception bias

- "**Availability Bias**": When animals are unavailable for detection.
- "**Perception Bias**": When observers fail to detect animals at distance 0 although they are available

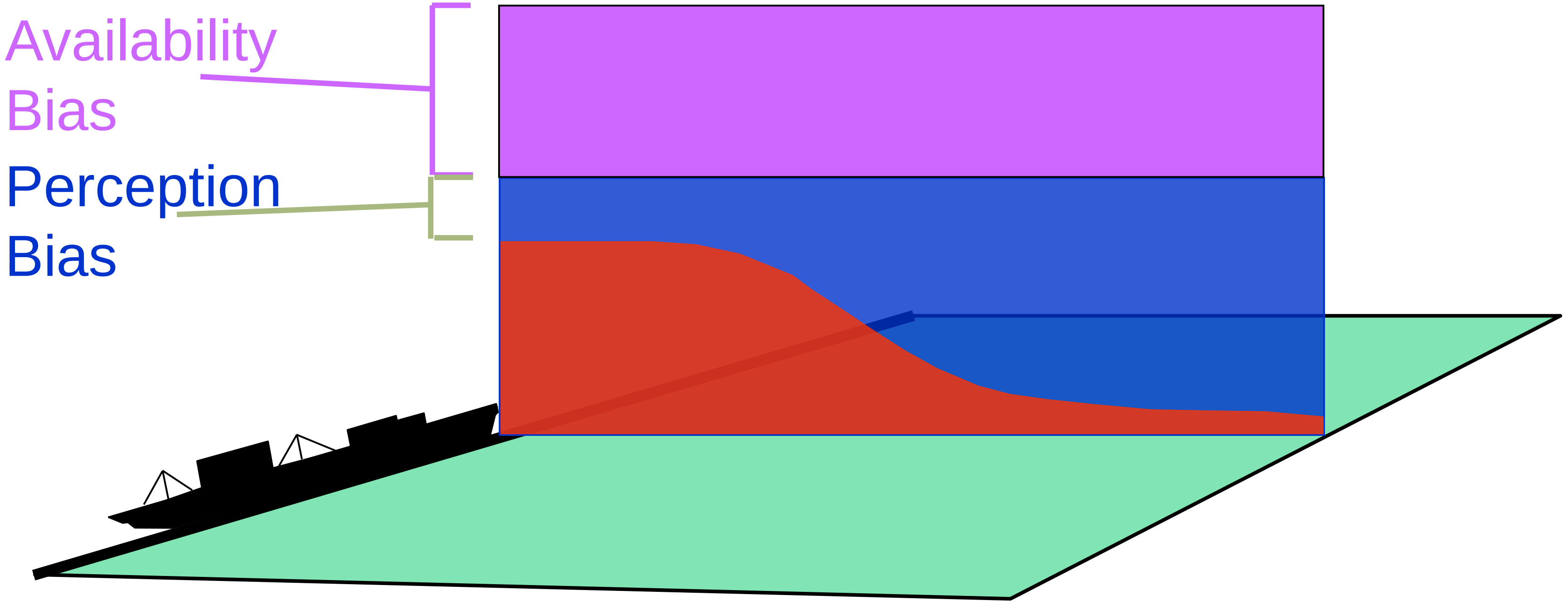


CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

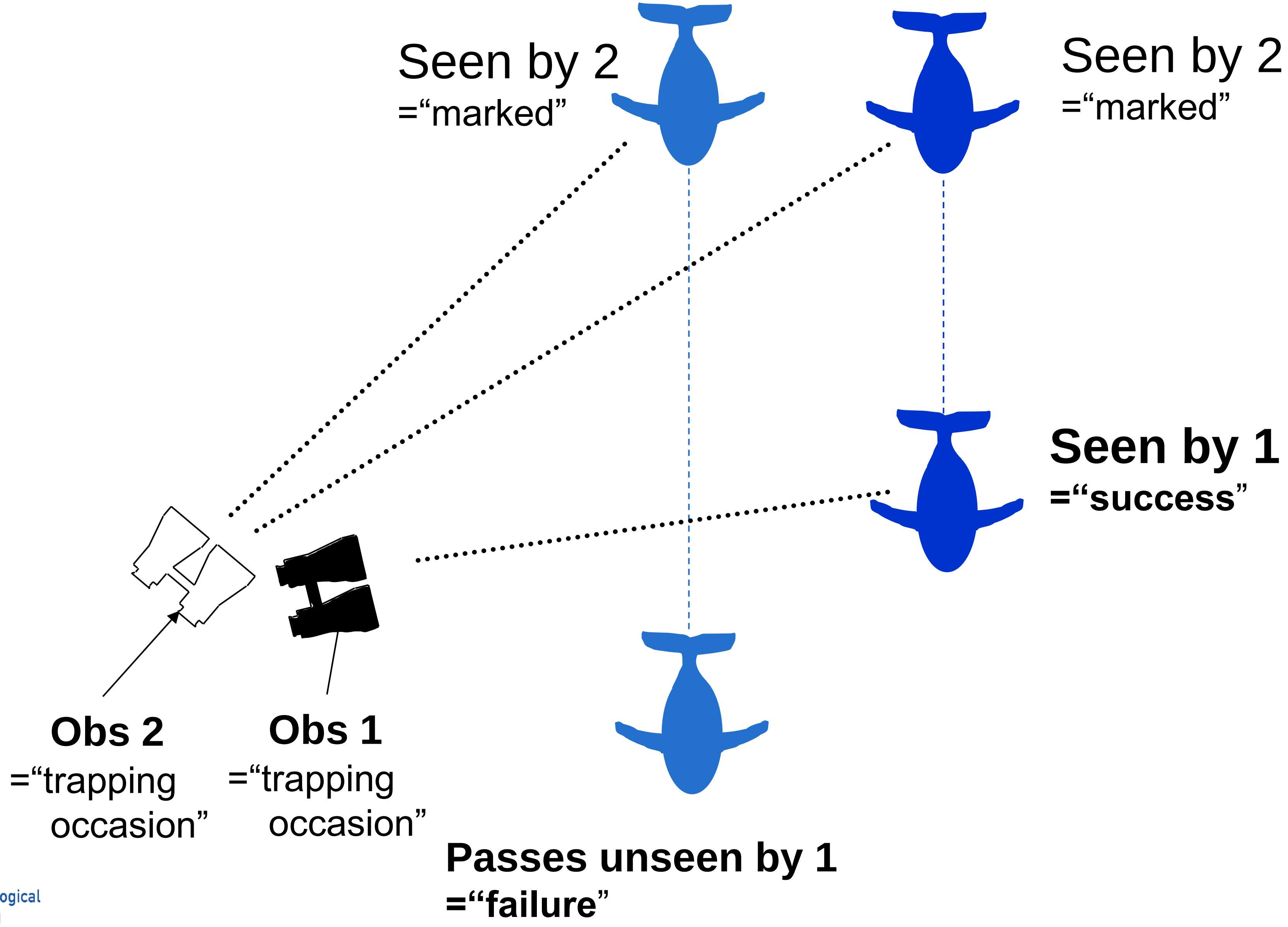
"**Availability Bias**": When animals are unavailable for detection.

"**Perception Bias**": When observers fail to detect animals on the transect although they are available



CREEM
Centre for Research into Ecological and Environmental Modelling
University of St Andrews

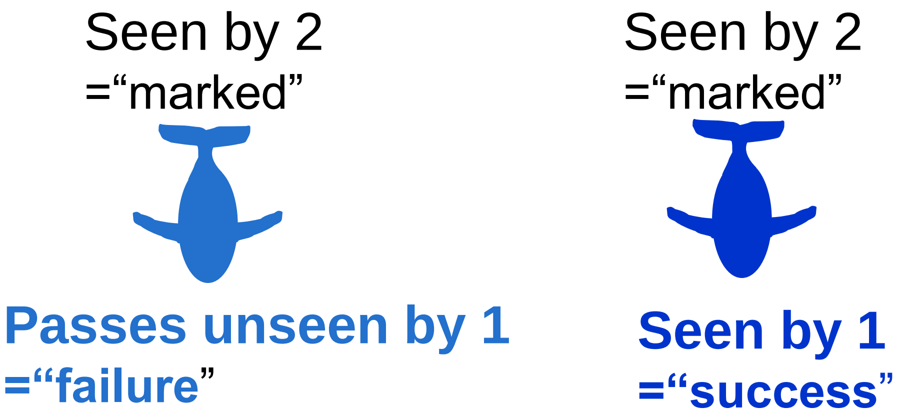
Visual Mark-Recapture



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Visual Mark-Recapture



- We know 2 animals passed (because Obs 2 saw them)
- Of these, Obs 1 saw 1
- So **estimate:**
  Pr(Obs 1 sees) = $\hat{p}_1 = \frac{1}{2} = \frac{n_{12}}{n_2}$ = $\frac{\text{number "duplicates"}}{\text{number seen by 2}}$

Simulated data example

Simulated 2000 animals between 0 and 1000m, with two observation platforms and detectability a function of distance from transect line and “visibility”.

$$n_2 = 831$$

$$n_{12} = 520$$

$$\hat{p}_1 = \frac{n_{12}}{n_2} = \frac{520}{831} = 0.626$$

$$n_1 = 835$$

$$\hat{N} = \frac{n_1}{\hat{p}_1} \times \frac{A}{2wL} = \frac{835}{0.626} \times 1 = 1334$$

Just in the surveyed transect strip

Why didn’t it work?

Unmodelled heterogeneity in detection probability!

Effect of heterogeneity - illustration

	N	p	$E(n_2)$	$E(n_{12})$	$E(n_1)$
Big animals	1000	0.9	900	810	900
Small animals	1000	0.1	100	10	100
Total	2000	0.5	1000	820	1000

Using the totals:

$$\hat{p}_1 = \frac{n_{12}}{n_2} = \frac{820}{1000} = 0.82 \qquad \hat{N} = \frac{n_1}{\hat{p}_1} = \frac{1000}{0.82} = 1220$$

Effect of heterogeneity - illustration

	N	p	$E(n_2)$	$E(n_{12})$	$E(n_1)$
Big animals	1000	0.9	900	810	900
Small animals	1000	0.1	100	10	100
Total	2000	0.5	1000	820	1000

Using the different types of animal:

$$\hat{p}_{1,Big} = \frac{n_{12,Big}}{n_{2,Big}} = \frac{810}{900} = 0.9 \qquad \hat{N}_{Big} = \frac{n_{1,Big}}{\hat{p}_{1,Big}} = \frac{900}{0.9} = 1000$$

$$\hat{p}_{1,Small} = \frac{n_{12,Small}}{n_{2,Small}} = \frac{10}{100} = 0.1 \qquad \hat{N}_{Small} = \frac{n_{1,Small}}{\hat{p}_{1,Small}} = \frac{100}{0.1} = 1000$$

$$\hat{N} = \hat{N}_{Big} + \hat{N}_{Small} = 1000 + 1000 = 2000$$

General formulation: $\hat{N} = \sum_{\text{groups } k} \frac{n_k}{\hat{p}_k} = \sum_{\text{individuals } i} \frac{1}{\hat{p}_i}$

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Effect of heterogeneity - conclusion

Unmodelled heterogeneity in detectability with mark-recapture type data causes

- Positive bias in estimation of p
- Negative bias in estimation of N

If you can model it correctly, the bias disappears

Sources of heterogeneity

Many! E.g.,

Animals
- Behaviour, Intrinsic visibility, Cluster size, Distance from the transect

Environment
- Habitat, Environmental conditions (mist, glare, sea state...)

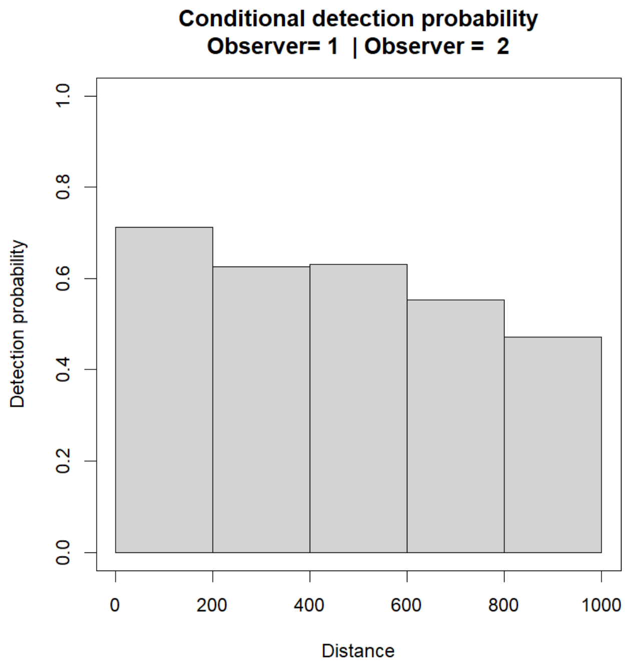
Observers
- Observer abilities, Observation platform (height, visibility, ...)

...

Simulated data example Revisited

Incorporate distance from transect line into the analysis

Distance	n_2	n_{12}	$\hat{p}_1$	n_1	$\hat{N}$
0-200m	281	200	0.711	283	398
200-400m	174	109	0.626	184	294
400-600m	149	94	0.631	144	228
600-800m	123	68	0.553	119	215
800-1000m	104	49	0.471	105	223



Summing across distance bands: $\hat{N}$ = 1358

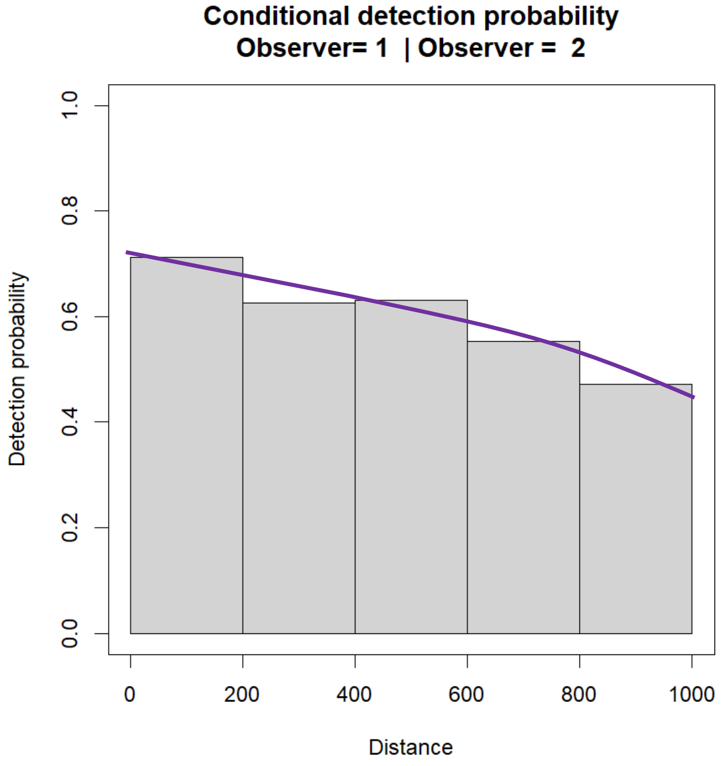
Bit better than previous estimate (1334)

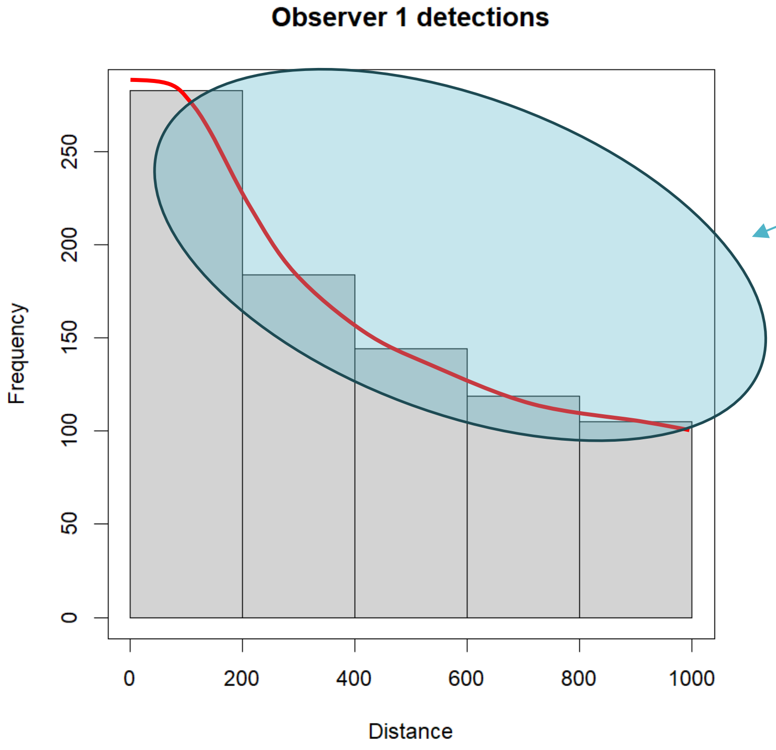
but not close to true value of 2000!

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Evidence still unmodelled heterogeneity

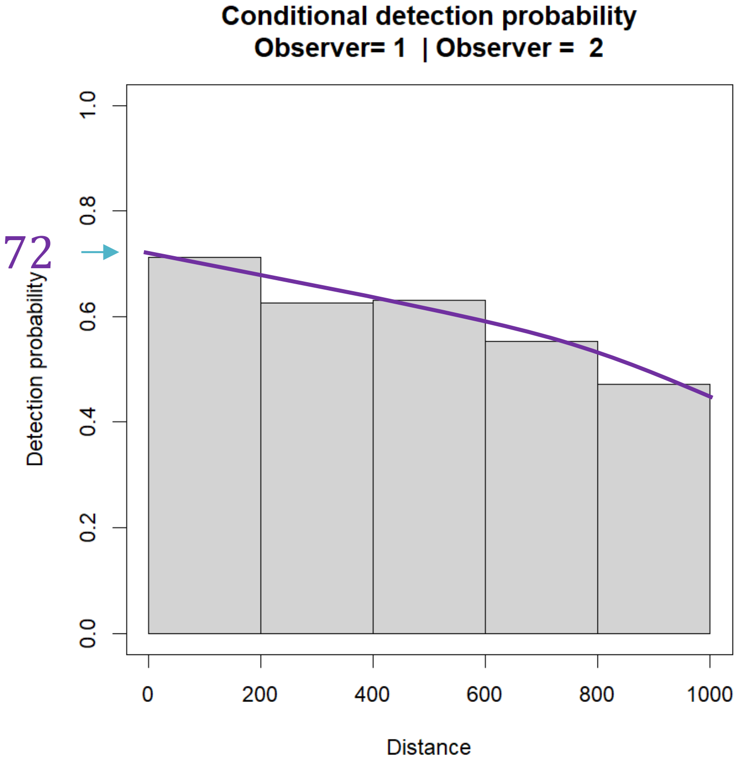


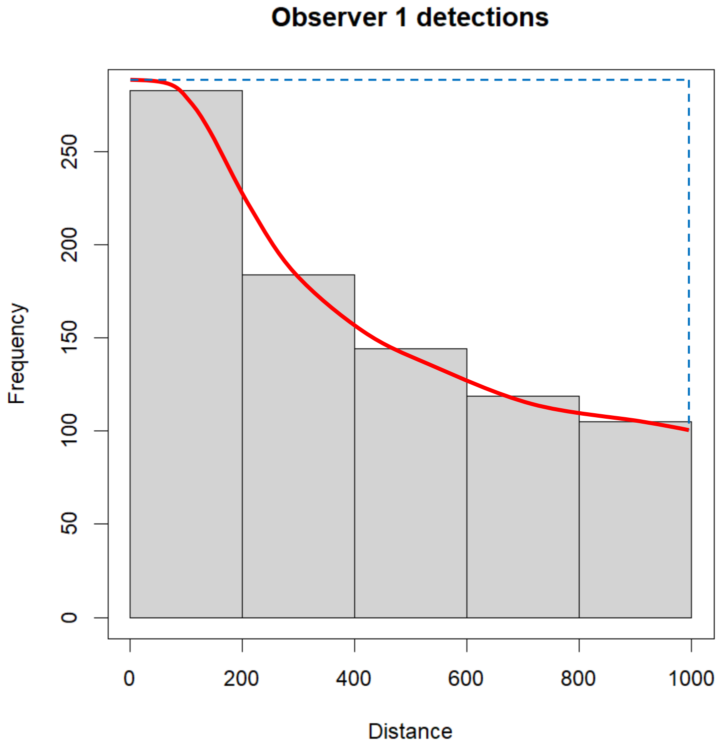


Unmodelled heterogeneity here

Dealing with unmodelled heterogeneity

$\hat{g}_1(0) \approx 0.72$





$$\hat{P}^*_{a,1} = \frac{\text{area under curve}}{\text{area under rectangle}}$$

$$\approx \frac{835}{283 \times 5} = 0.59$$

$$\hat{N} = \frac{n_1}{\hat{g}_1(0) \times \hat{P}^*_{a,1}} = \frac{835}{0.72 \times 0.59} = 1966$$

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Full vs point independence models

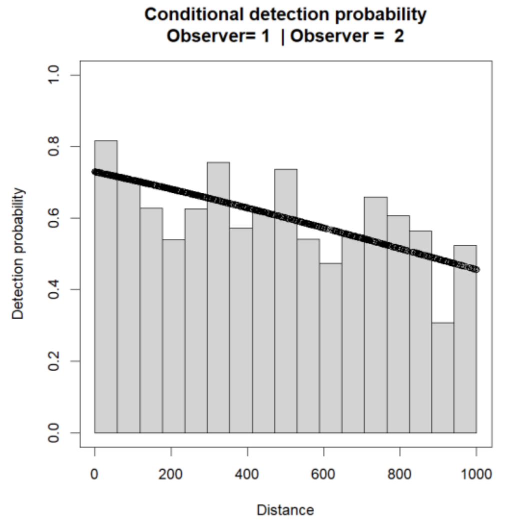
Full independence model

Uses detections from one observer as "trials" to obtain detection probability for the other

Detection function model is a binary regression with logit link function (a.k.a. logistic regression) – *"mark-recapture model"*

Assumes probability of detection by the observer setting up the trial is independent of the probability of detection by the other observer at ***all distances***, given covariates – "*full independence*"

```
mrmodel = ~glm(~distance)
```



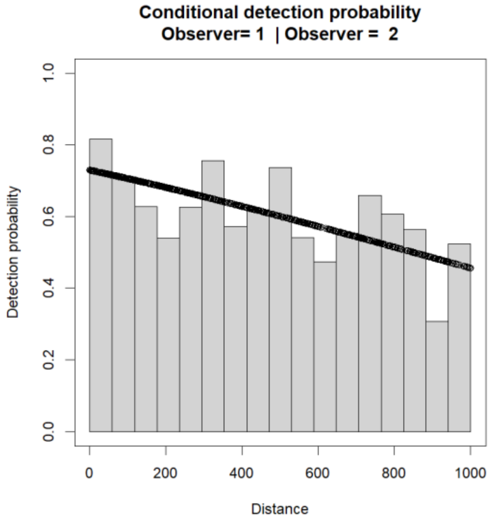
	Estimate	SE	CV
Average p	0.5978057	0.01834089	0.03068034
Average primary p(0)	0.7293052	0.02213859	0.03035572
N in covered region	1396.7747929	52.02831290	0.03724889

Full vs point independence models

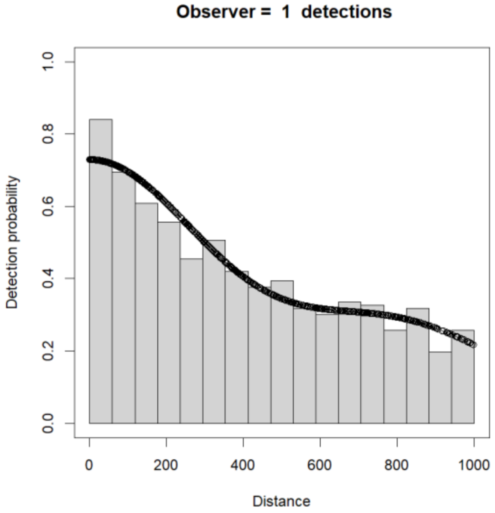
Point independence model

- Uses ***mark-recapture model*** to get $g(0)$ (called $p(0)$ in some literature)
- Uses standard ***distance sampling model*** to get $\hat{P}_a^*$
- Combines them to estimate overall average detection prob
- Assumes probability of detection by the observer setting up the trial is independent of the probability of detection by the other observer at ***0 distance only***, given covariates – "***point independence***"

```
mrmodel =
  ~glm(~distance)
```




```
dsmodel =
  ~mcds(key = "hn",
  adj.series = "cos",
```




```
                          Estimate           SE         CV
Average p                0.4210556   0.02526106 0.05999459
N in covered region   1983.1109662 129.93081451 0.06551868
```

Full vs point independence models

Full independence model

Sensitive to unmodelled heterogeneity – negative bias.

Assumption of uniform animal distribution not required – so useful for responsive movement.

Don't use unless you have to!

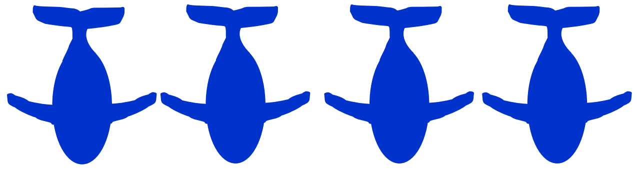
Point independence model

Less sensitive to unmodelled heterogeneity.

Assumption of uniform animal distribution required for ds model – so no good if there is responsive movement.

Use unless there is responsive movement.

Simulated data example
Model selection

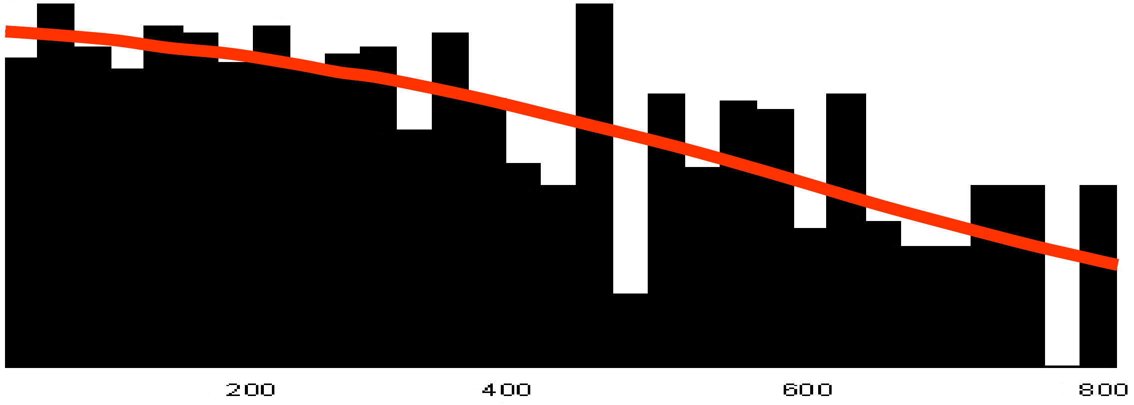


Model type	MR model	DS model	AIC	$\hat{N}$
Full independence	~1	-	12636.83	1334
Full independence	~distance		12552.17	1396
Point independence	~distance	~hn + cos(2)	12506.41	1983
Full independence	~distance + visibility	-	12430.23	1540
Full independence	~distance x visibility	-	12269.08	1917

General class of models are known as “Mark-Recapture Distance Sampling” (MRDS)

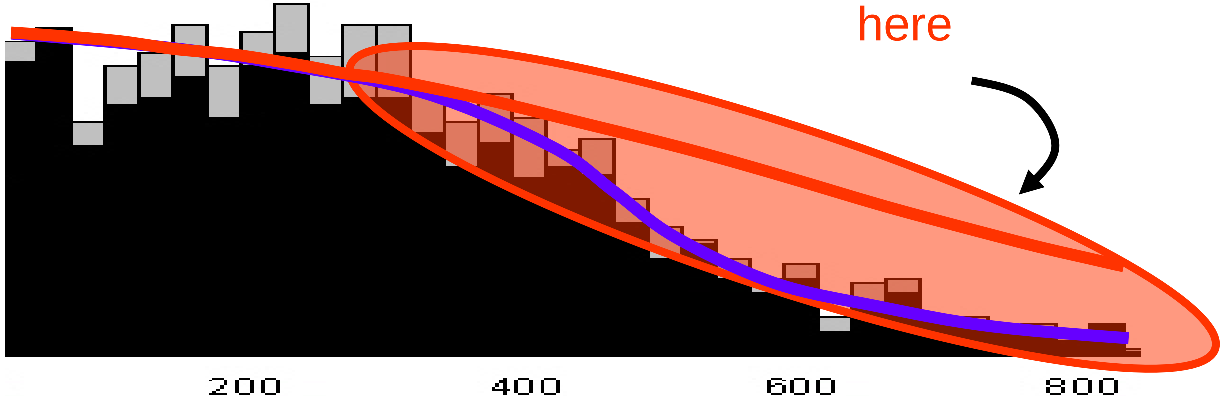
Real data example: pack-ice seals

Proportion of Observer 2 detections seen by Observer 1



Observer 1 detections

Unmodelled Heterogeneity
here



CREEM
Centre for Research into Ecological and Environmental Modelling

Data from Southwell et al. (2007)

University of St Andrews

Configuration: Trial

Observer 2

↓

sets up trials for

↓

Observer 1

↓

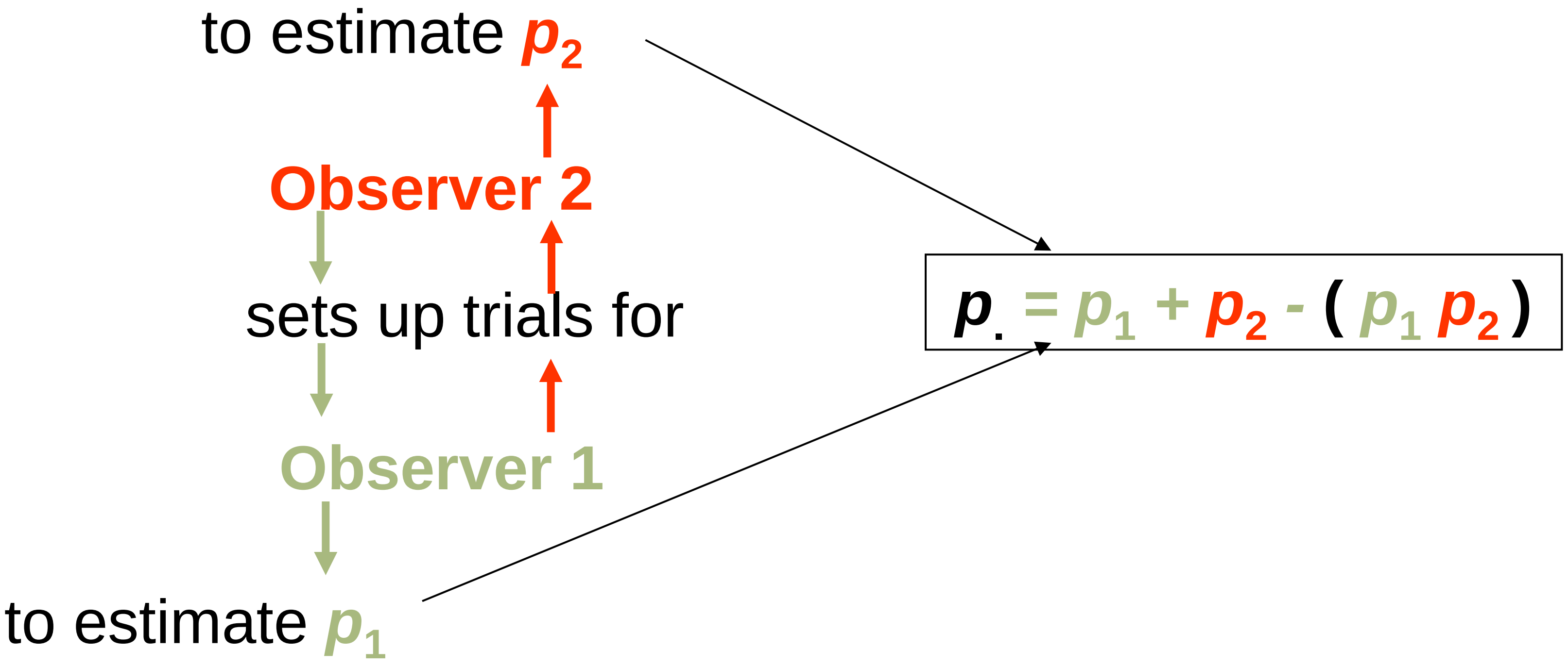
to estimate p_1

The Observer at the end of an arrow must be independent of
the Observer at the start of the arrow

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Configuration: Independent Observer



The Observer at the end of an arrow must be independent of
the Observer at the start of the arrow

CREEM
Centre for Research into Ecological and Environmental Modelling
University of St Andrews

Abundance estimation

Trial $$\hat{N} = \sum_{seen\ by\ 1} \frac{1}{\hat{p}_1(x_i, \ldots)}$$

Independent Observer $$\hat{N} = \sum_{seen} \frac{1}{\hat{p}_.(x_i, \ldots)}$$

Comparing configurations

Trial	Independent observer
Only requires observer 1 to be isolated from observer 2 (who sets up trials).	Requires both observation platforms to be isolated from one another.
Can be robust to responsive movement if observer 2 searches far ahead and their perpendicular distances are the ones used for analysis.	Not applicable, as both observers' set up trials, and it is generally better if they search different distances ahead (reduces availability bias).
Uses less data – only trials from observer 1.	Uses more data – trials from both observers.

Critical assumptions of MRDS

We have the required level of independence between observers

- Trial configuration: one-way independence – observer 1 independent of observer 2
- Independent observer configuration: two-way independence

No unmodelled heterogeneity

- Full independence models: at all distances
- Point independence models: at zero distance

Duplicates (resightings) are known

Duplicate identification

Can use a dedicated "duplicate identifier"

- Or for trial configuration, observer 2 (or one observer on that team) can track animals until they go abeam

Record positions and times of sightings as precisely as possible

- Allows rule-based duplicate identification after the survey

Record ancillary data – behaviour, etc.

Can record measure of confidence in duplicate identification

- Allows analysis using different levels of confidence

Related MRDS models not covered

Limiting independence

- Further relaxes assumption about unmodelled heterogeneity – assumes heterogeneity tends to zero as probability of detection approaches 1
- No standard software
- Buckland et al. (2009)

Point transects

- Implemented in standard software
- Laake et al. (2011)

Summary & Conclusions

In standard methods we assume g(0)=1

But g(0) can be <1 because of availability or perception bias

One approach to combat this is to deploy two (semi-) independent observation platforms, and identify duplicate detections

These data can be analyzed using Mark Recapture Distance Sampling (MRDS) models

Results are sensitive to unmodelled heterogeneity

- Collect relevant covariates
- Consider point- or full-independence models

Thought: given the complications, can you make g(0) close to 1 by altering your field methods?

References

Borchers, D. L., Laake, J. L., Southwell, C., & Paxton, C. G. M. (2006). Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics*, *62*, 372–378. https://doi.org/10.1111/j.1541-0420.2005.00493.x

Buckland, S. T., Laake, J., & Borchers, D. L. (2010). Double observer line tansect methods: Levels of independence. *Biometrics*, *66*, 169–177. https://doi.org/10.1111/j.1541-0420.2009.01239.x

Burt, M. L., Borchers, D. L., Jenkins, K. J., & Marques, T. A. (2014). Using mark-recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution*, *5*, 1180–1191. https://doi.org/10.1111/2041-210X.12294

Laake, J. L., & Borchers, D. L. (2004). Methods for incomplete detection at distance zero. In S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, & L. Thomas (Eds.), *Advanced distance sampling* (pp. 127–208). Oxford: Oxford University Press.

Laake, J. L., Collier, B. A., Morrison, M. L., & Wilkins, R. N. (2011). Point-based mark-recapture distance sampling. *Journal of Agricultural, Biological and Environmental Statistics*, *16*, 389–408. https://doi.org/doi.org/10.1007/s13253-011-0059-5

Southwell, C., Borchers, D., de la Mare, B., Paxton, C. G. M., Burt, L., & de la Mare, W. (2007). Estimation of detection probability on aerial surveys of Antarctic pack-ice seals. *Journal of Agricultural, Biological, and Environmental Statistics*, *12*, 1–14. https://doi.org/10.1198/108571107X162920

.