Making Distance Sampling Work

- Assumptions and effect of violation
- Reliable distance sampling
- Pooling robustness
- Examples of imperfect data

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Recap of distance sampling

There are two stages to estimating abundance

Stage 1: given n, how many objects are in the surveyed/covered region (of size a), N_a

Need to estimate P_a *(or* $f(0)$ *or ESW, etc.)*

$$\hat{N}_a = {}^{n}\!/_{\hat{P}_a}$$

Stage 2: given $\hat{N}_a$, how many objects are in study region (of size A), N

'Scale up' from what we see in the survey region to the whole study region

$$\hat{N} = \frac{\hat{N}_a}{{}^{a}\!/_{A}}$$

Assumptions for estimating N_a (stage 1)

1. **Animals distributed independently of line or point**

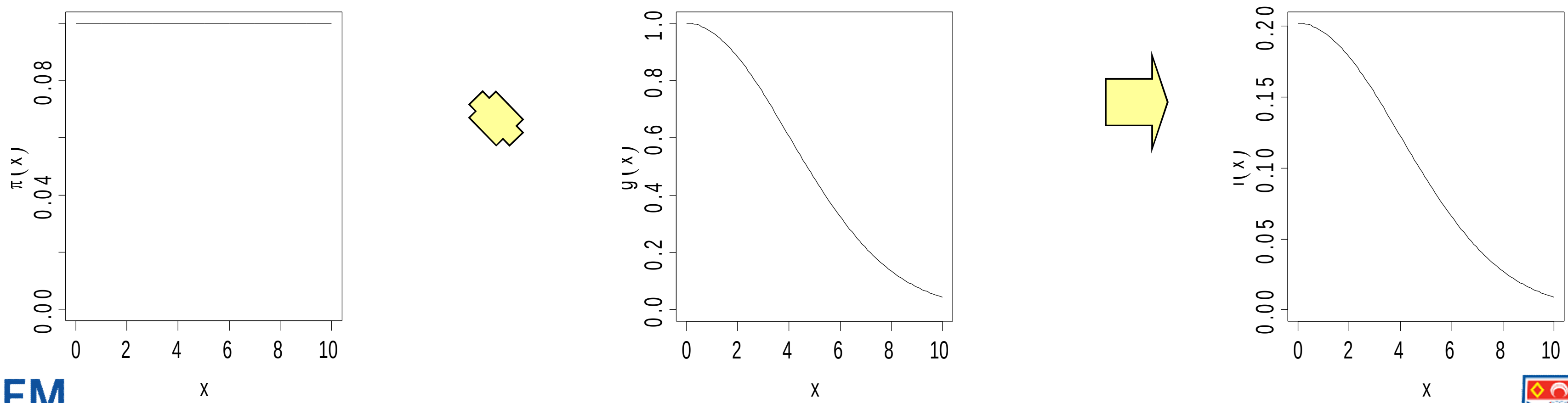
   This ensures the true distribution of animals with respect to the line or point is known

   Violated by non-random line/point placement

   Substantial violation can produce substantial bias (e.g. roadside counts)

e.g. for line transects

True distribution of animals | Detection function, $g(x)$ | Observed distribution, $f(x)$



Assumptions for estimating N_a (stage 1)

2. **All animals on the line or point are detected i.e. $g(0)=1$**

   It is a critical assumption - violation causes negative bias

   e.g. if $g(0)=0.8$, estimates of N are 80% of true N on average

Images courtesy of FreeDigitalPhotos.net

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Assumptions for estimating N_a (stage 1)

3. Observation process is a 'snapshot'

Other ways to phrase this:

Observers are moving much faster than the animals

Animals do not move before they can be detected

Problems of independent/non-responsive movement

An animal moving independently of the observer (compared to moving in response to the observer) produces positive bias; size of bias depends on relative rate of movement of observer and animal, and type of survey.

Point transect methods in particular need to use 'snapshot' method.

Assumptions for estimating N_a (stage 1)

3. Observation process is a 'snapshot' (continued…)

Problems of responsive movement

Responsive movement can cause large bias

It can occur ***within*** *a single line/point or* ***between*** *lines/points*

If animals are 'driven' from one line/point to the next ahead of the observer, positive bias will result.

Note: movement independent of observer outwith 'snapshot' is fine – in this case, the same animal can be detected on multiple lines/transects

Assumptions for estimating N_a (stage 1)

4. **Distances are measured accurately**

   Random errors cause bias.

   Bias is generally small for line transect estimators,

   Can be large for point transect estimators.

   Both are sensitive to systematic bias and to rounding to 0 distance (or angle).

   Can use grouped data collection.

5. **Detections are independent**

   Violation has little effect. (Model selection methods for $g(x)$, such as AIC, are somewhat affected)

Assumptions for estimating *N* given N_a (stage 2)

1. Lines or points are located according to a survey design with appropriate randomization

    We use properties of the survey design to extrapolate from the surveyed/covered region to the study region ('design-based')

    Non-random survey design means density in surveyed/covered region may not be representative of density in study region. Also variance may be biased.

Image courtesy of FreeDigitalPhotos.net

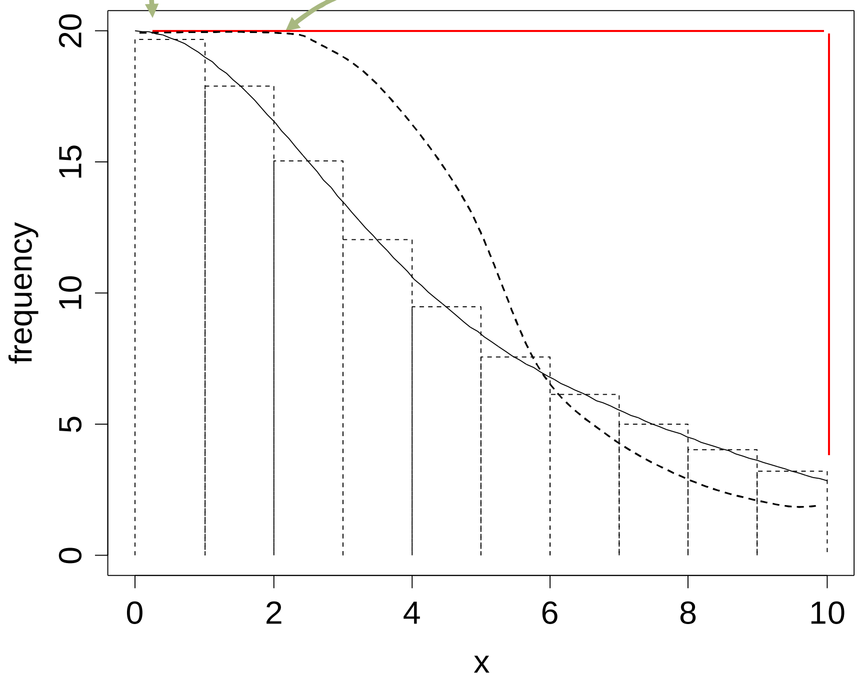
Reliable distance sampling (1)

1. Reliable estimation of P_a (or $f(0)$ or ESW, etc.)

In addition to the assumptions, we would like:

SHAPE CRITERION
Detection function should have a 'shoulder' (i.e. $g'(0)=0$)

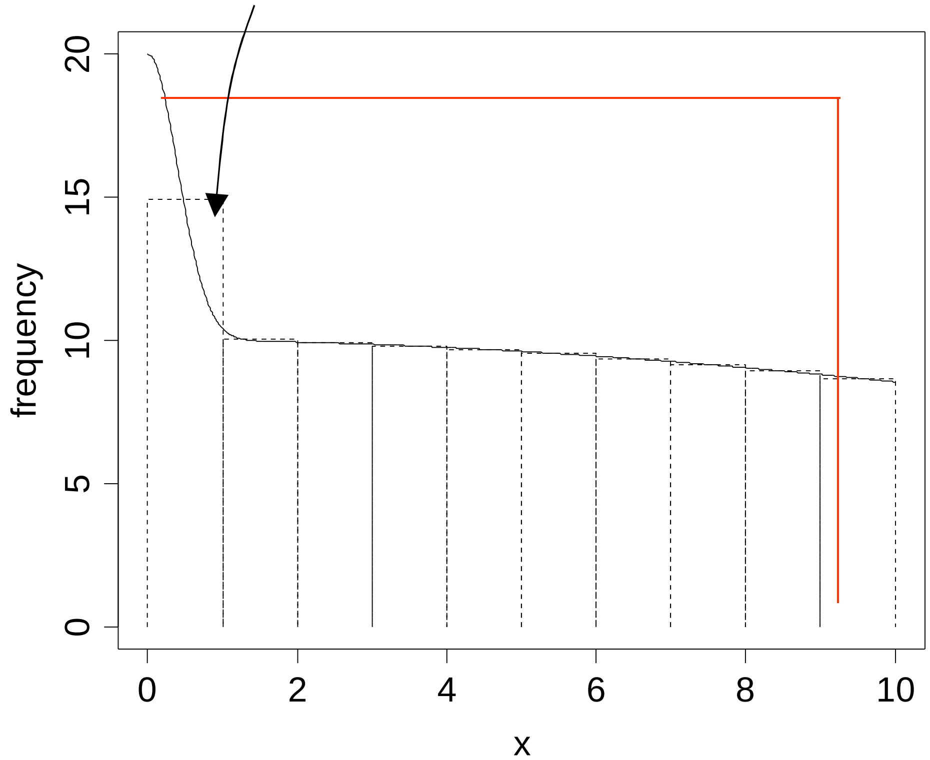
Data that have a wide shoulder are preferable



A wide shoulder makes it easier to estimate area under rectangle (or $f(0)$, etc.)

(1) Reliable estimation of P_a

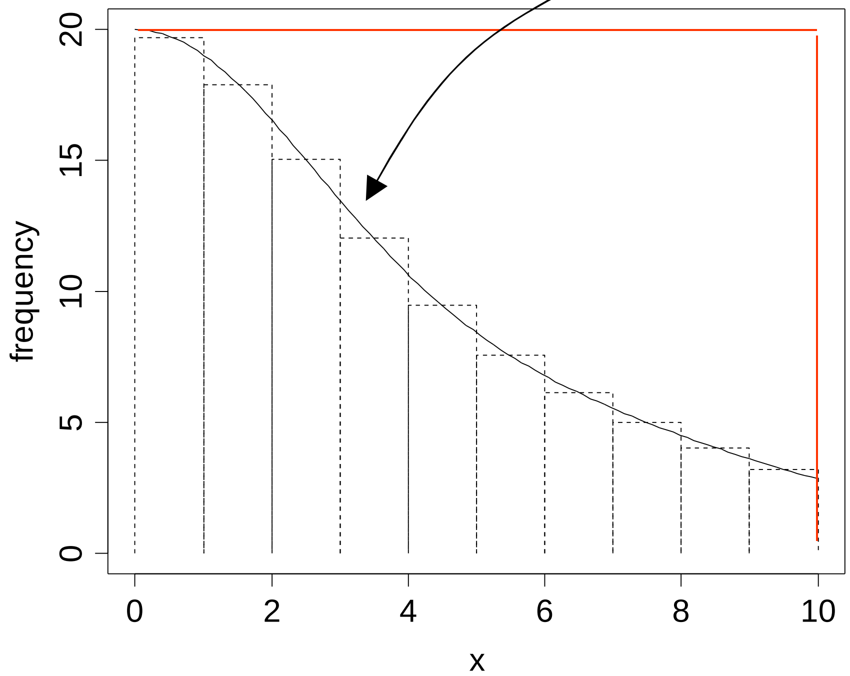
Good field methods will avoid a ‘spike’ like this



Avoid a) rounding distances (and angles) to zero,
b) ‘guarding the trackline’

(1) Reliable estimation of P_a (cont.)

Flexible detection function model can fit the data (see later)



Sample size of observations (~60-80)

- less for detection functions with 'easy' shapes
- more for point transects and 'difficult shapes'.

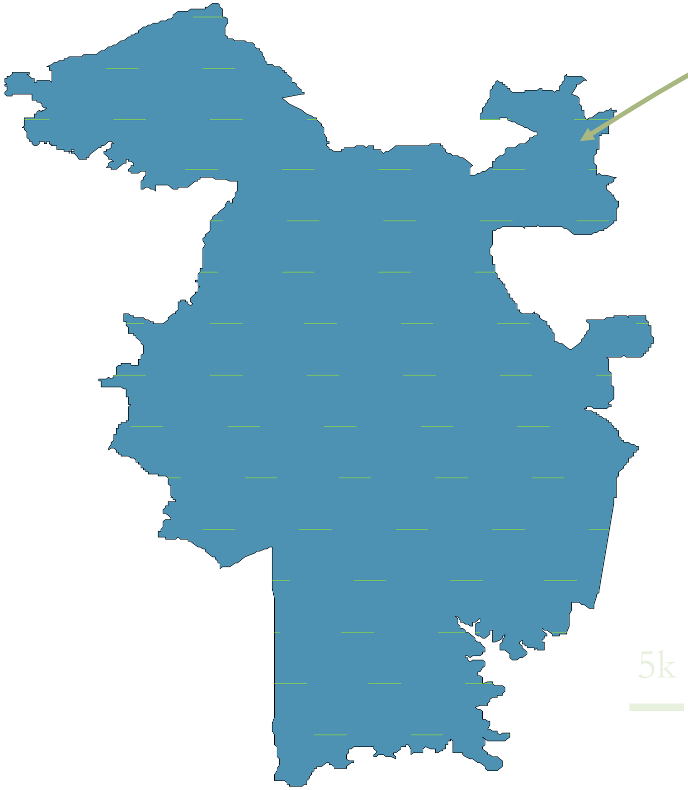
Reliable distance sampling (2)

2. Reliable estimation of N from N_a

In addition to the assumption of randomized design, we would like a 'large' sample of lines or points (20 or more), evenly distributed through the study region

see lecture on survey design

e.g. surveys of tiger prey in India



Photos: Ullas Karanth

CREEM
Centre for Research into Ecological and Environmental Modelling

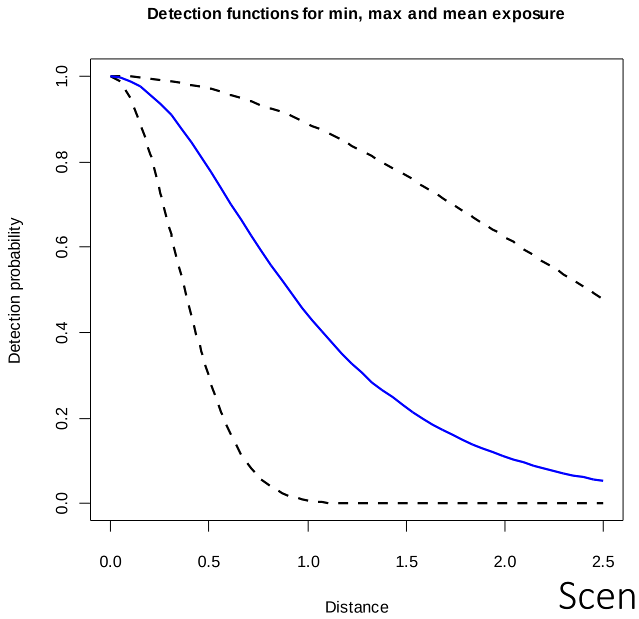
University of St Andrews

Pooling robustness

Individuals can have quite different detection functions, but this produces little bias (up to a point!)

'Pooling robustness' = robust to pooling of multiple detection functions

e.g. Simulation study (unpublished!) Truth = 1000 animals

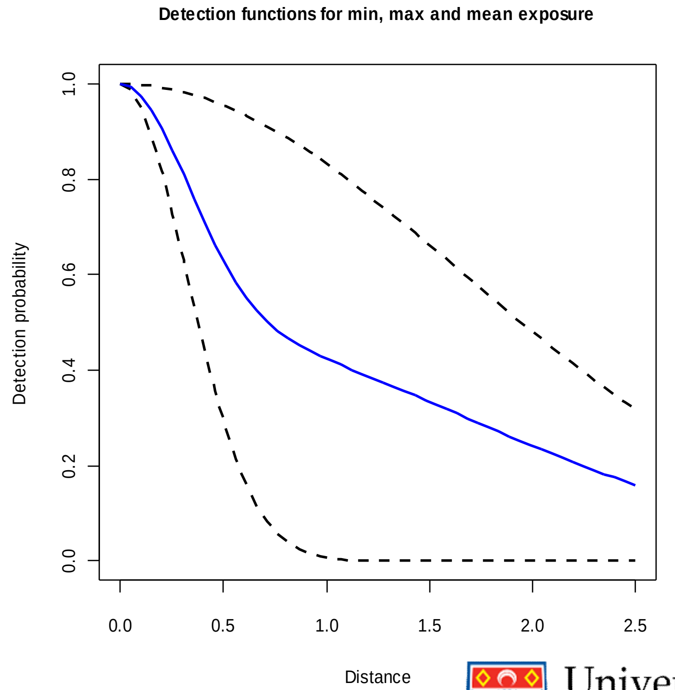


Scenario 1: animals have a gamma distribution of detection functions between min and max shown.

Mean estimate from simulation: 984 animals (SE 2.3). Bias -1.6%

Scenario 2: half of animals have max detection function, half have minimum.

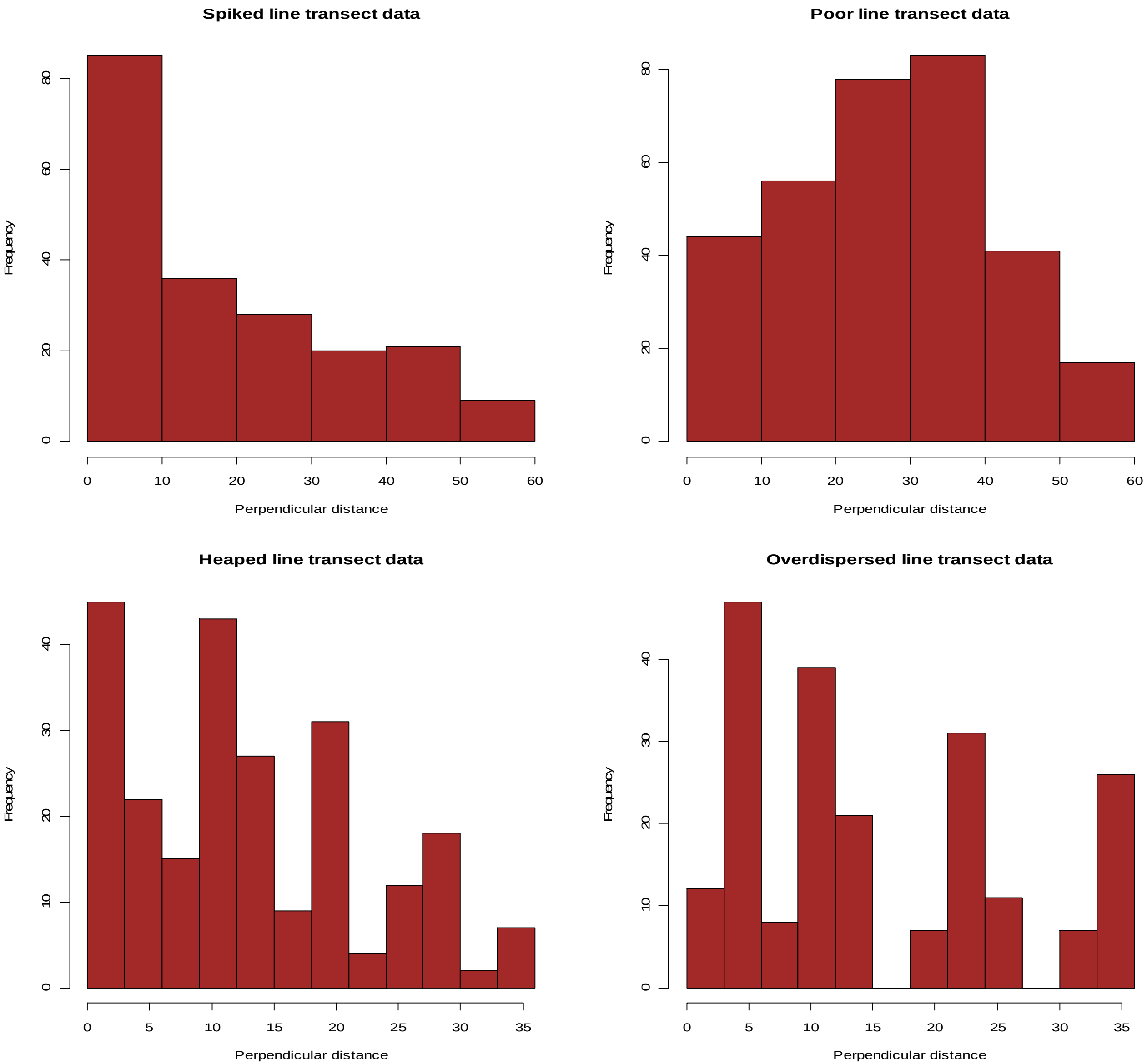
Mean estimate from simulation: 976 animals (SE 2.7). Bias -2.4%



CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Non-ideal data



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews