Assessment of model performance

- Likelihood
- AIC
- Absolute measures of model fit
  - Chi-squared test
  - Q-Q plots
  - Kolmogorov-Smirnov and Cramér-von Mises tests

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

Likelihood

$f(x)$ = probability density function of x

$f(x)\,dx$ = Pr (animal was between x and $x+dx$ from the line,
given it was detected between 0 and w) for small dx

When distances are exact, the likelihood is given by

$$L = \prod_{i=1}^{n} f(x_i) = f(x_1) \times f(x_2) \times \ldots \times f(x_n)$$

x_i = distance of i^{th} detected animal from the line.

We fit $f(x)$ by finding the values for the parameters of $f(x)$ (or equivalently $g(x)$) that maximize L (or $\log_e(L)$).

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Akaike's Information Criterion

$$\text{AIC} = -2\log_e(L) + 2q$$

L is the maximized likelihood (evaluated at the maximum likelihood estimates of the model parameters)

and q is the number of parameters in the model.

- Select the model with smallest AIC
- Gives a relative measure of fit

Limitations of AIC

Cannot be used to select between models when:

- sample size n differs
- truncation distance w differs
- data are grouped, and cut points differ
- data are grouped in one analysis and ungrouped in the other

Goodness-of-Fit

- Chi-squared test for grouped (interval) data
  - if data are exact, we must specify interval cut points to perform the test
- Q-Q plots and related tests for exact data

Chi-squared tests

Define u distance intervals, with n_i detections in interval i, i = 1, ..., u.

Then
$$\chi^2 = \sum_{i=1}^{u} \frac{(n_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i}$$

where $n = \sum_i n_i$

and $\hat{\pi}_i$ is the proportion of the area under the estimated pdf, $\hat{f}(x)$, that lies in interval i.

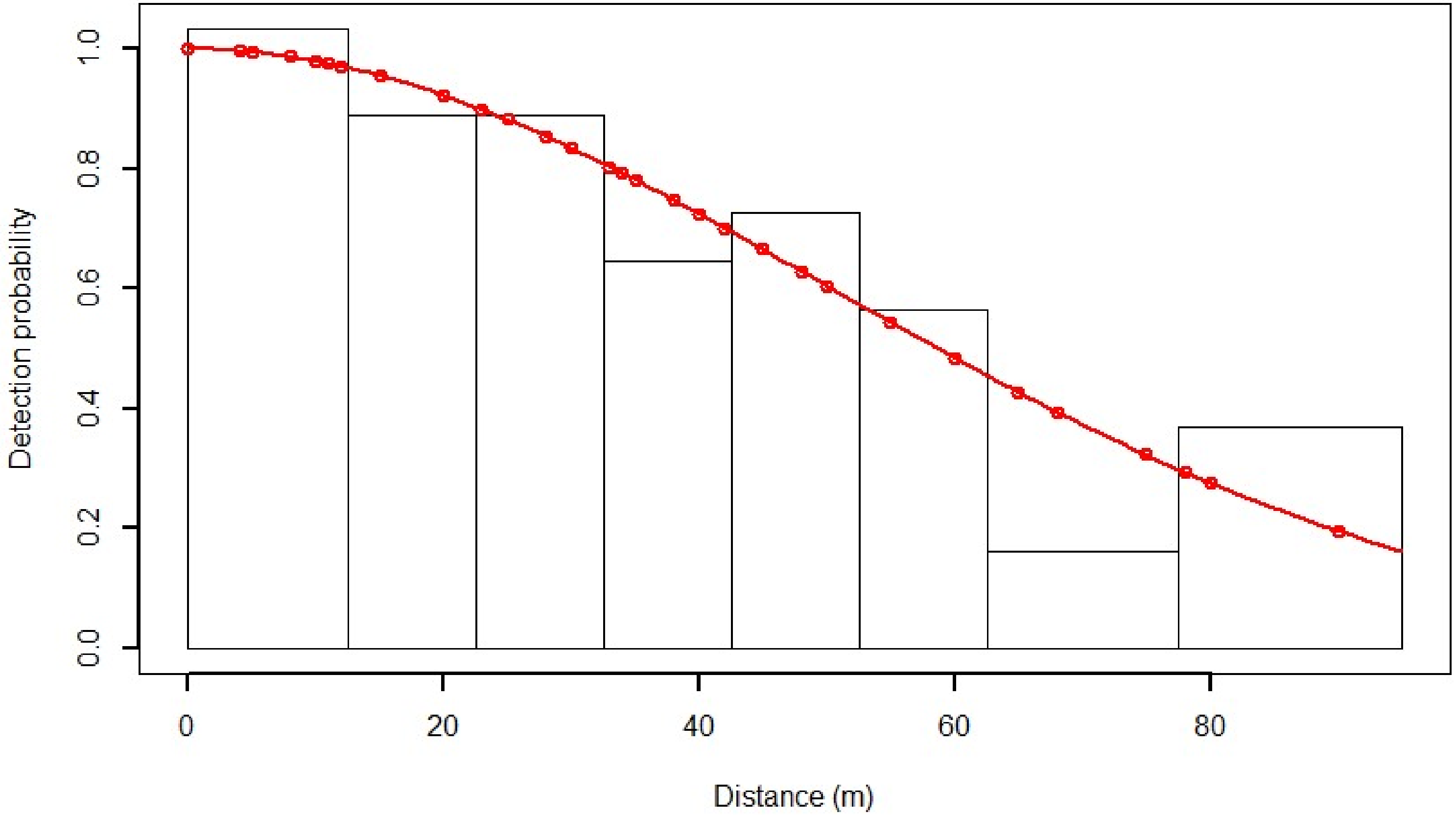
If the model is 'correct': $\chi^2 \sim \chi^2_{u-q-1}$

q = no. of parameters

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

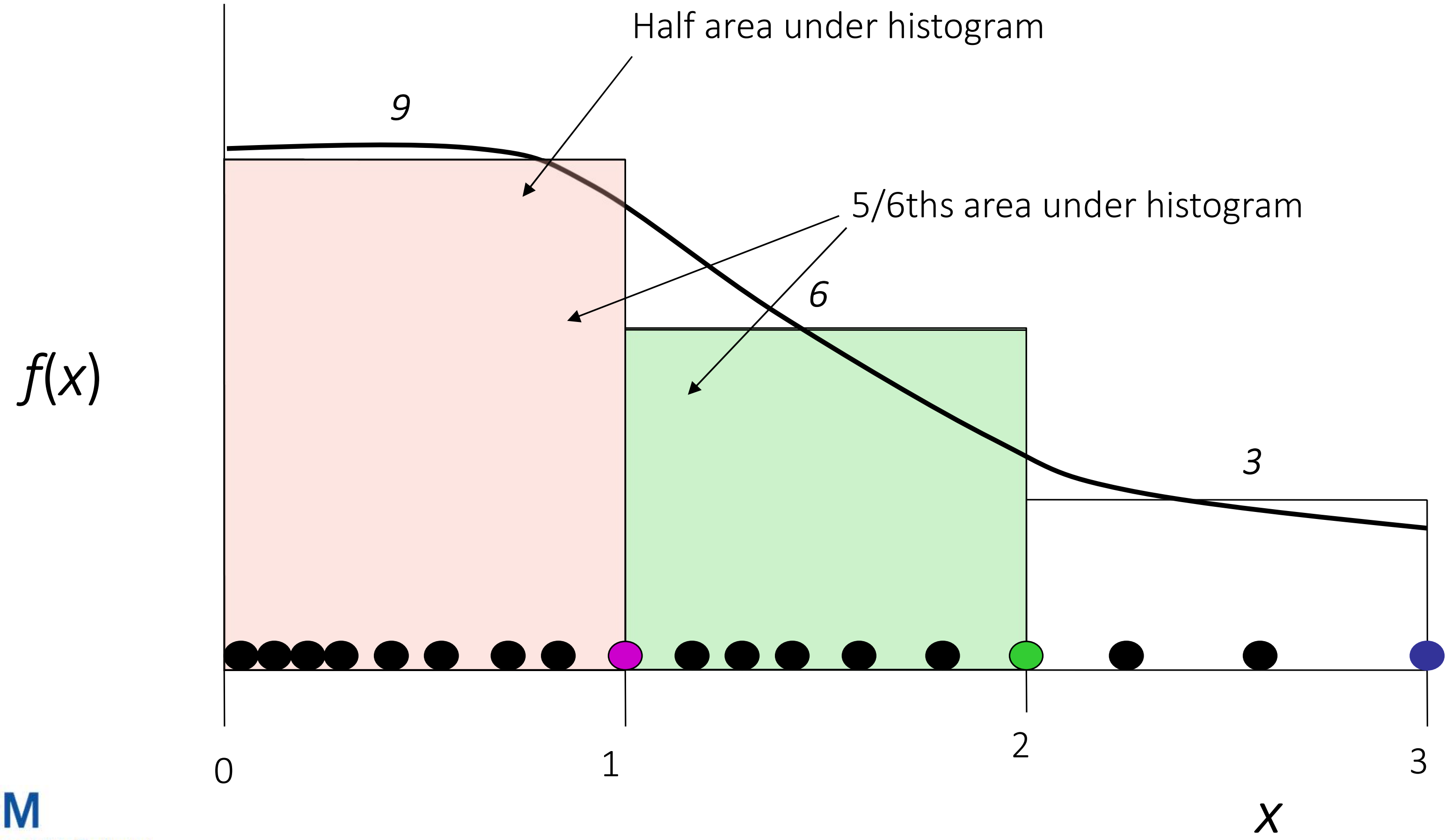
Chaffinch line transect data



χ^2 goodness-of-fit test

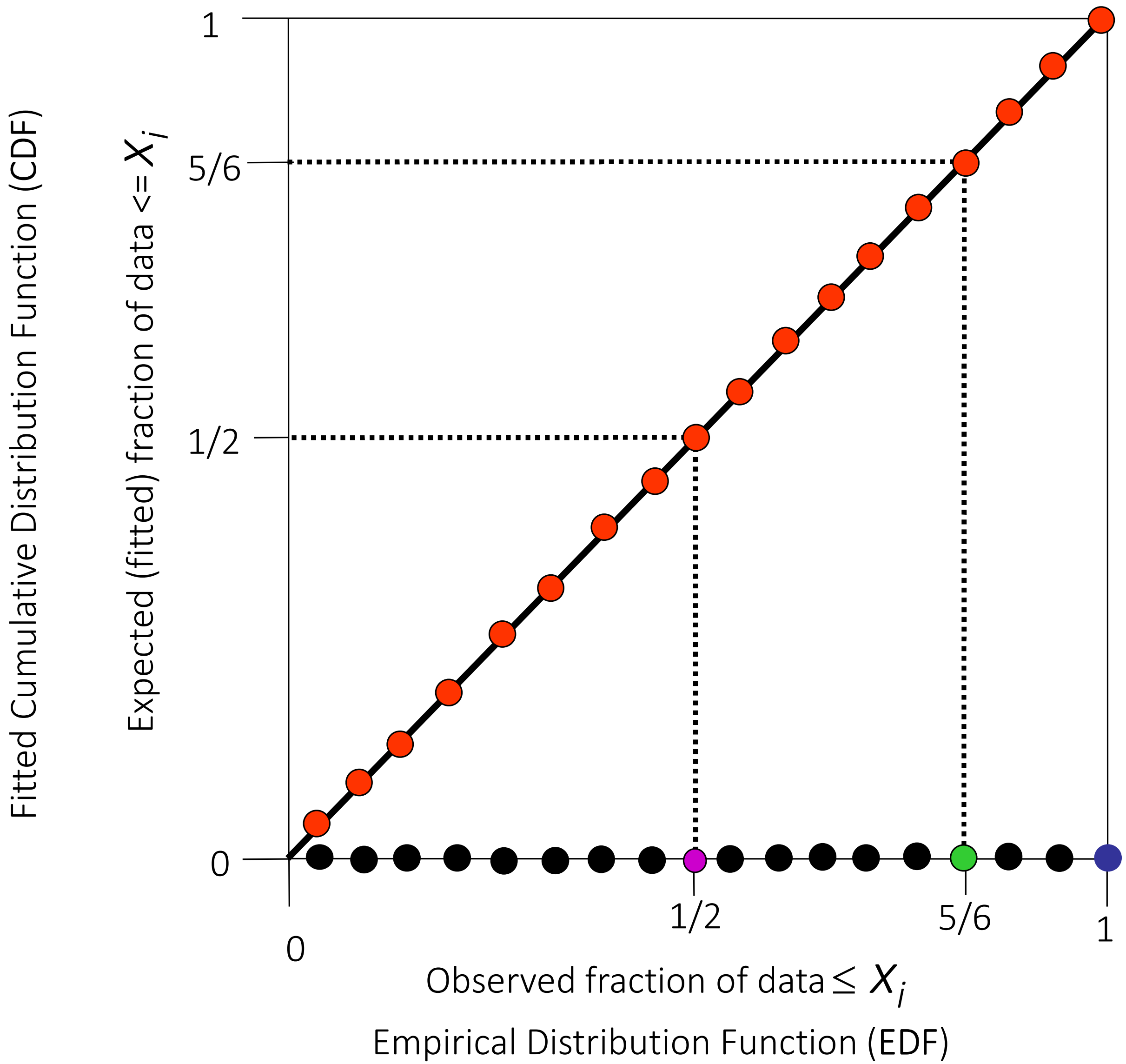
```
Goodness of fit results for ddf object

Chi-square tests
             [0,12.5] (12.5,22.5] (22.5,32.5] (32.5,42.5]
Observed  16.00000000 11.00000000   11.000000   8.0000000
Expected  15.31832030 11.62653282   10.623975   9.3264854
Chisquare  0.03033539  0.03376272    0.013309   0.1886631
          (42.5,52.5] (52.5,62.5] (62.5,77.5] (77.5,95]     Total
Observed    9.0000000  7.00000000    3.000000  8.000000 73.000000
Expected    7.8658030  6.37326777    6.960224  4.905391 73.000000
Chisquare   0.1635437  0.06163138    2.253286  1.952261  4.696791

P = 0.58325 with 6 degrees of freedom
```

Q-Q Plots and Related Tests



CREEM
Centre for Research into Ecological and Environmental Modelling

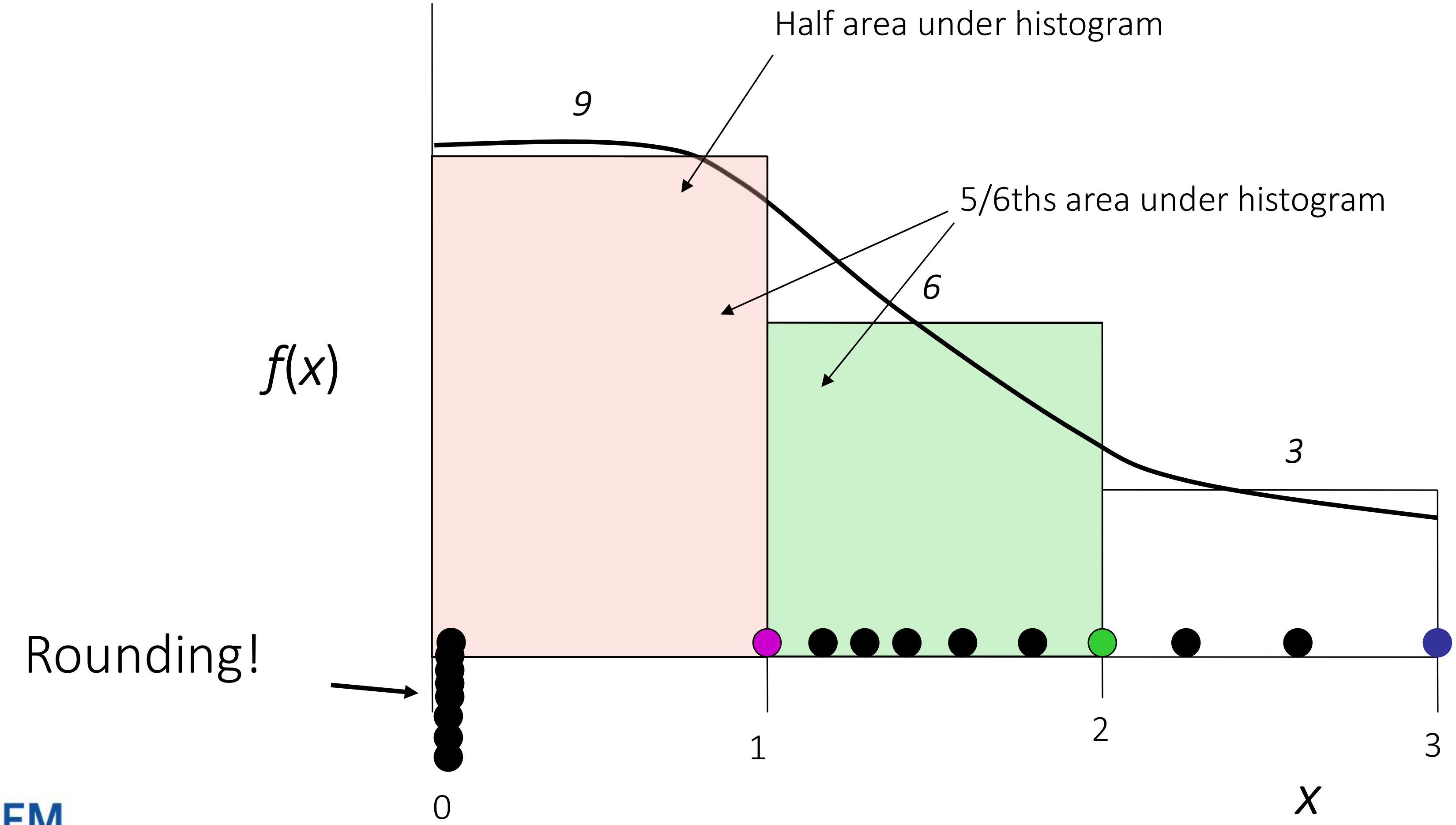
University of St Andrews



CREEM
Centre for Research into Ecological and Environmental Modelling

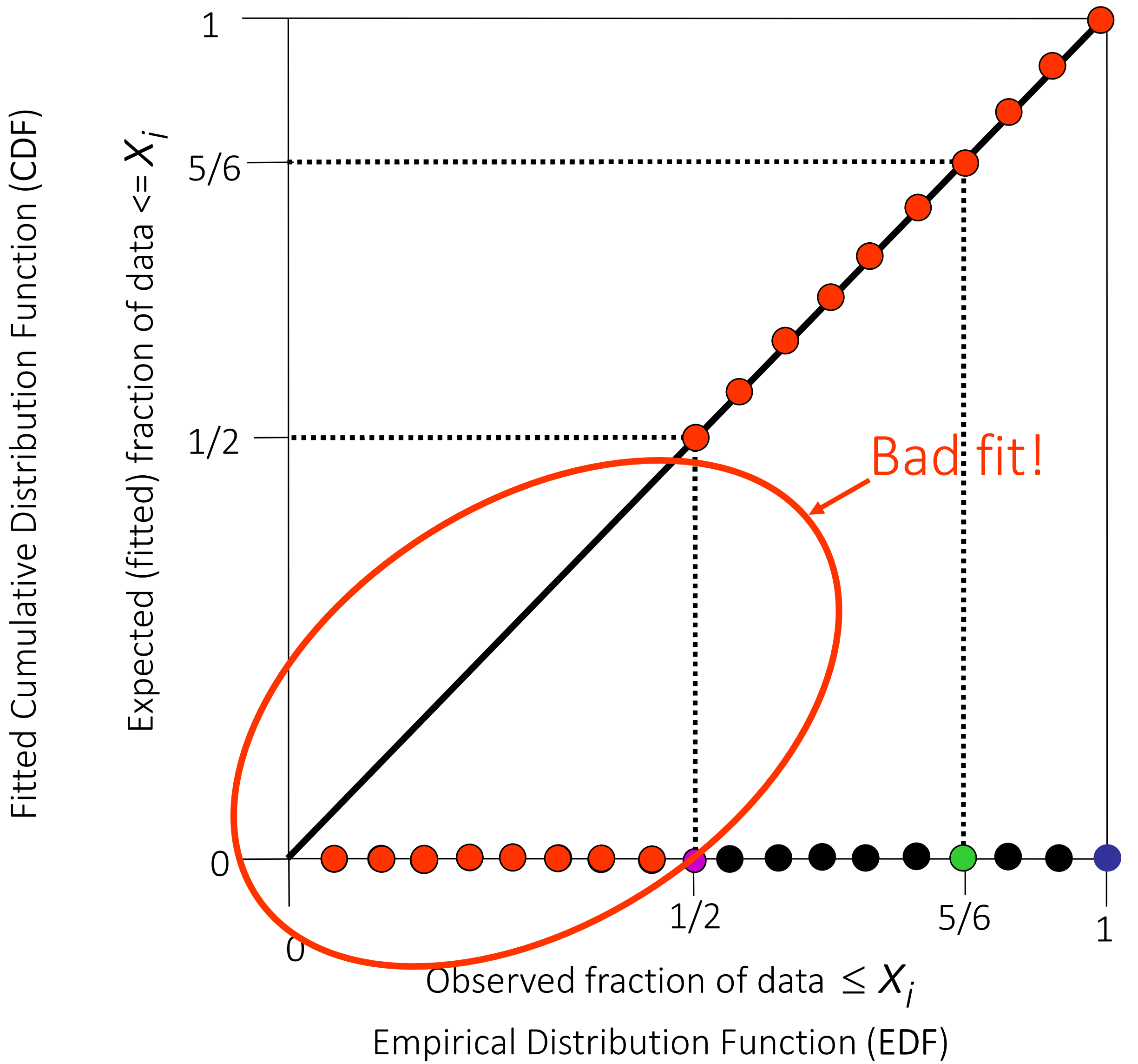
University of St Andrews

Example: Rounding to zero



CREEM Centre for Research into Ecological and Environmental Modelling

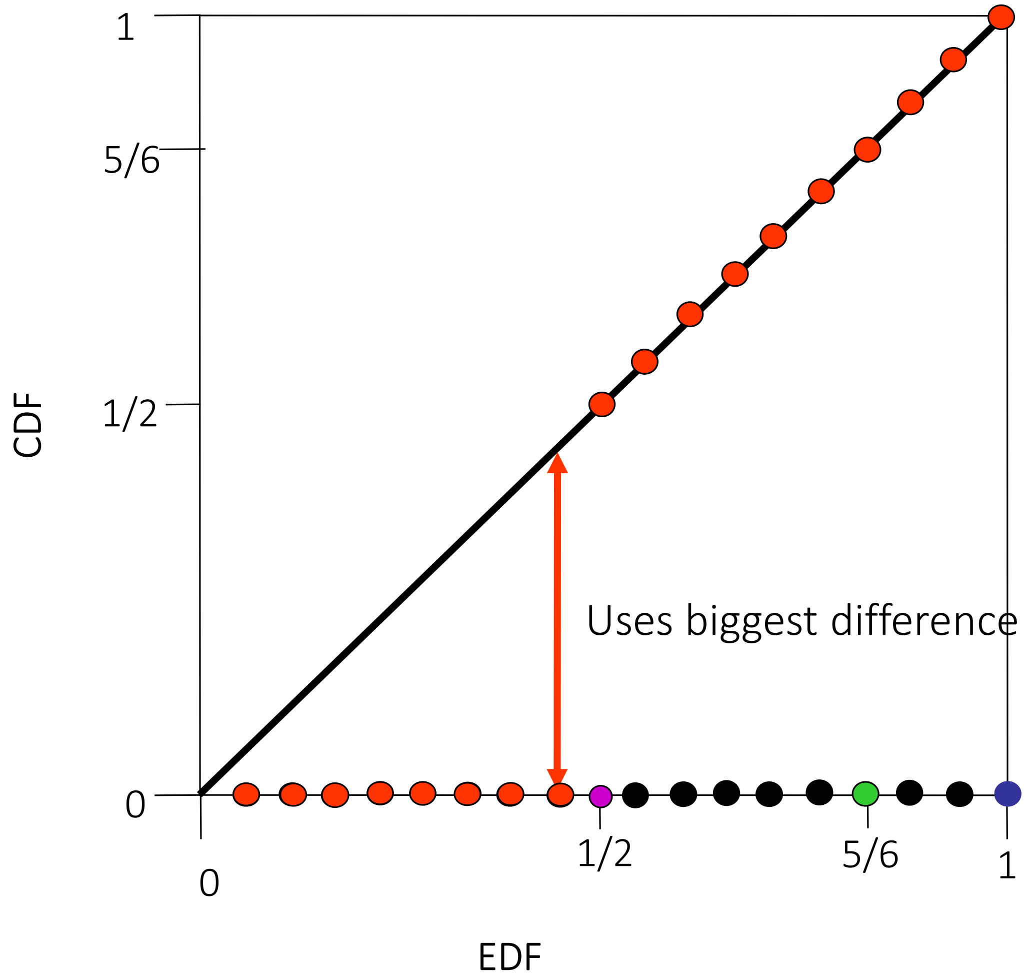
University of St Andrews



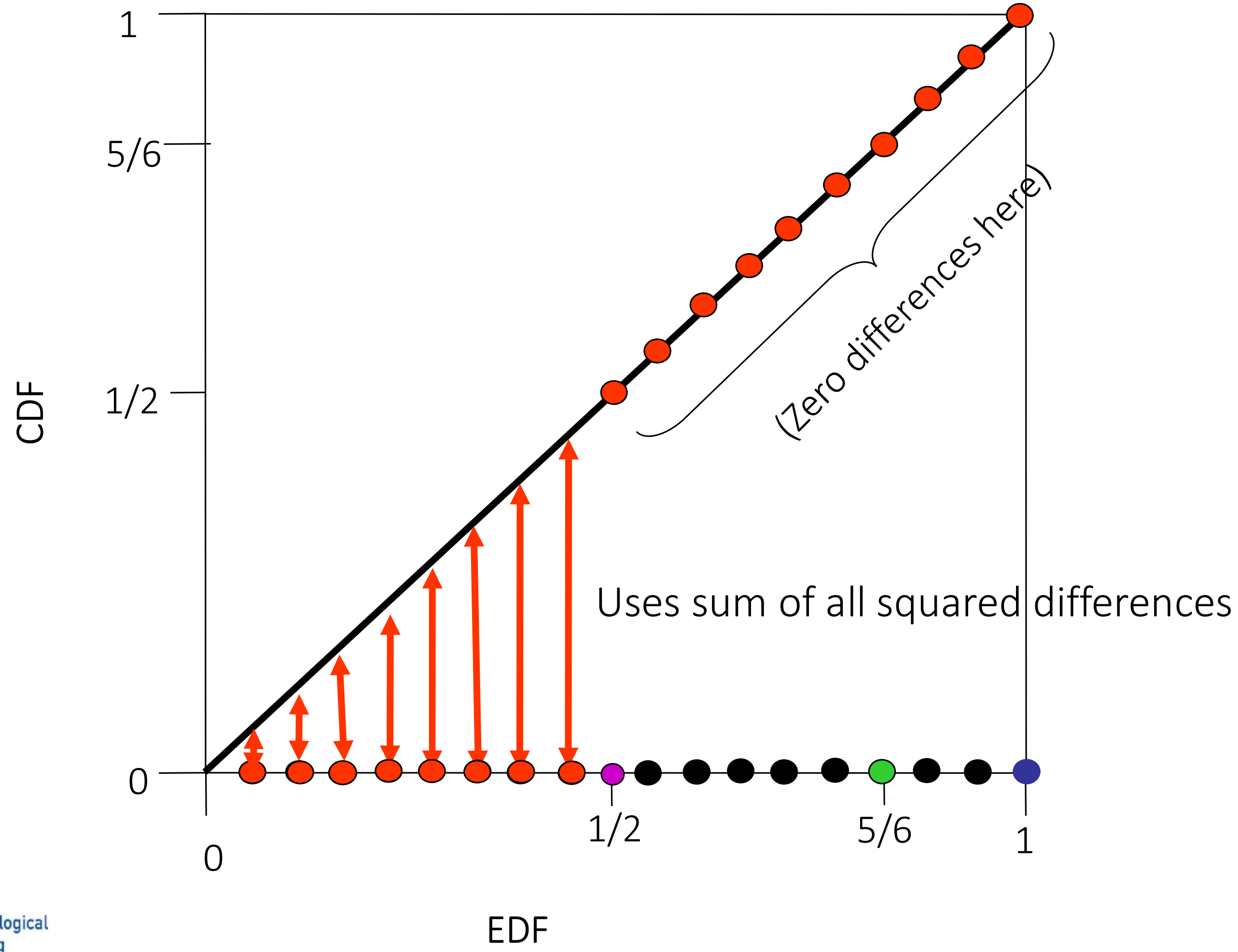
CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Kolmogorov-Smirnov test



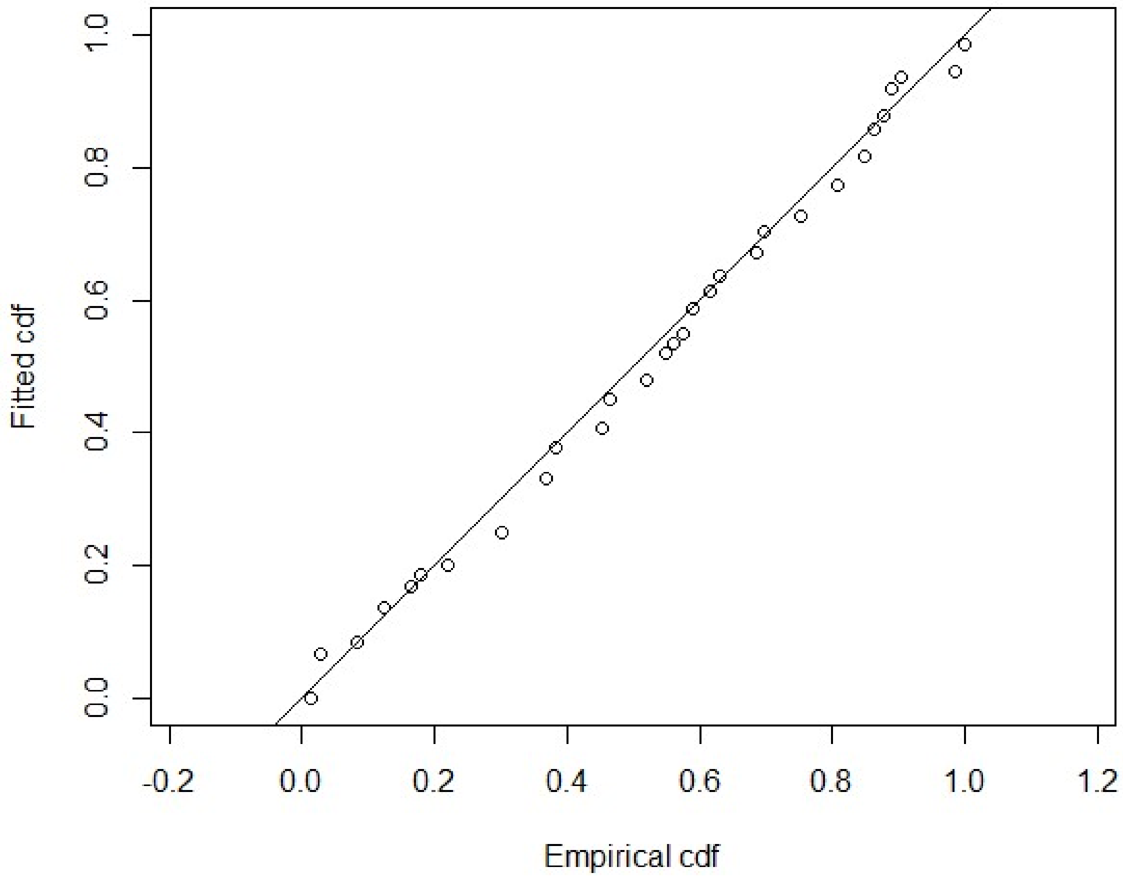
Cramér-von Mises test



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Chaffinch line transect Q-Q plot



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

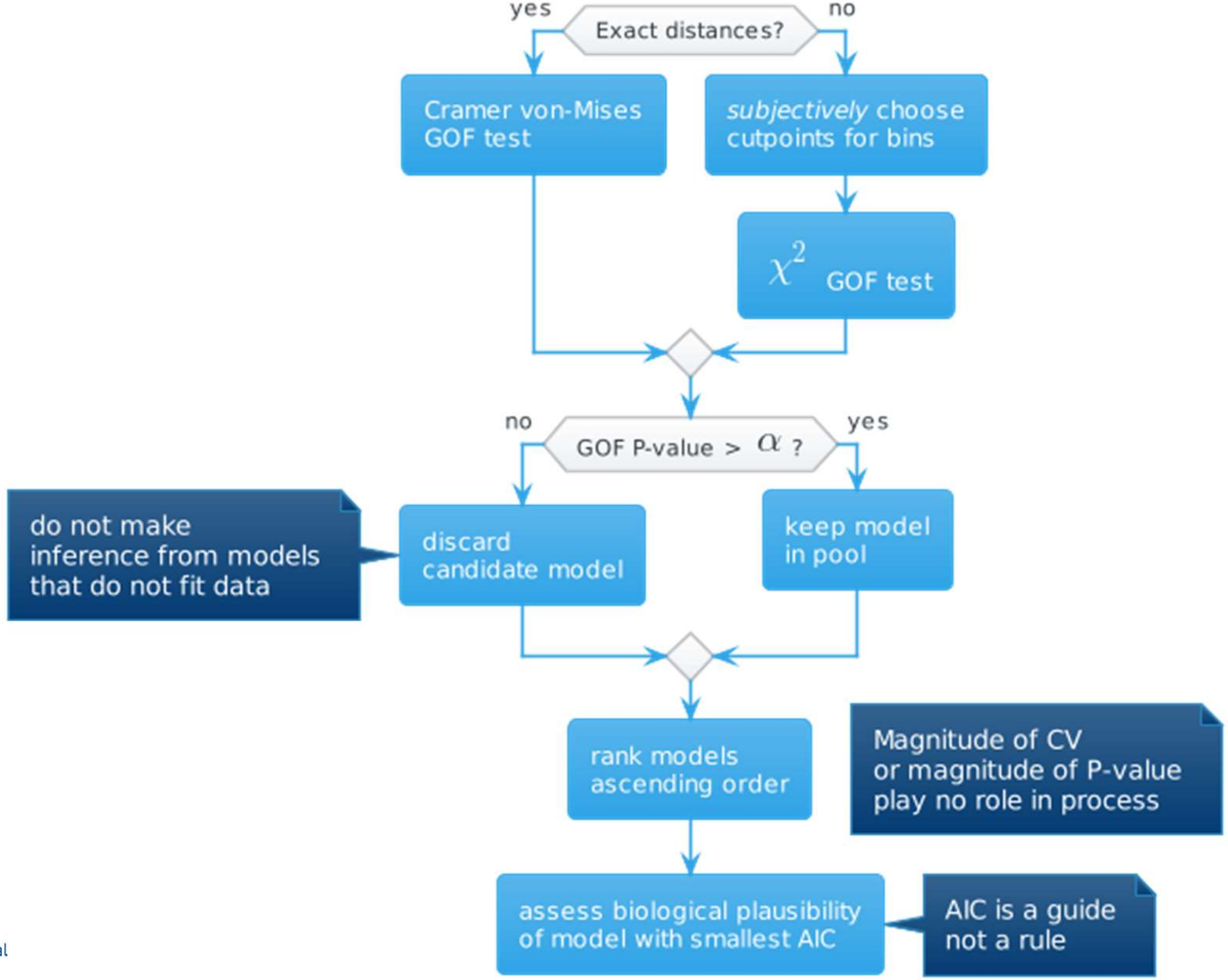
K-S test and Cramer-von Mises test

```
Distance sampling Kolmogorov-Smirnov test
Test statistic = 0.0572767 p-value = 1
 (p-value calculated from 100/100 bootstraps)

Distance sampling Cramer-von Mises test (unweighted)
Test statistic = 0.0367951 p-value = 0.948916
```

Q-Q Plot Summary

- Q-Q plots show goodness-of-fit at "high resolution" – without requiring grouping into intervals
- Kolmogorov-Smirnov test and Cramér-von Mises test are goodness-of-fit tests that do not require grouping

Tools of model selection



CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Making Distance Sampling Work

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

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

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

CREEM Centre for Research into Ecological and Environmental Modelling

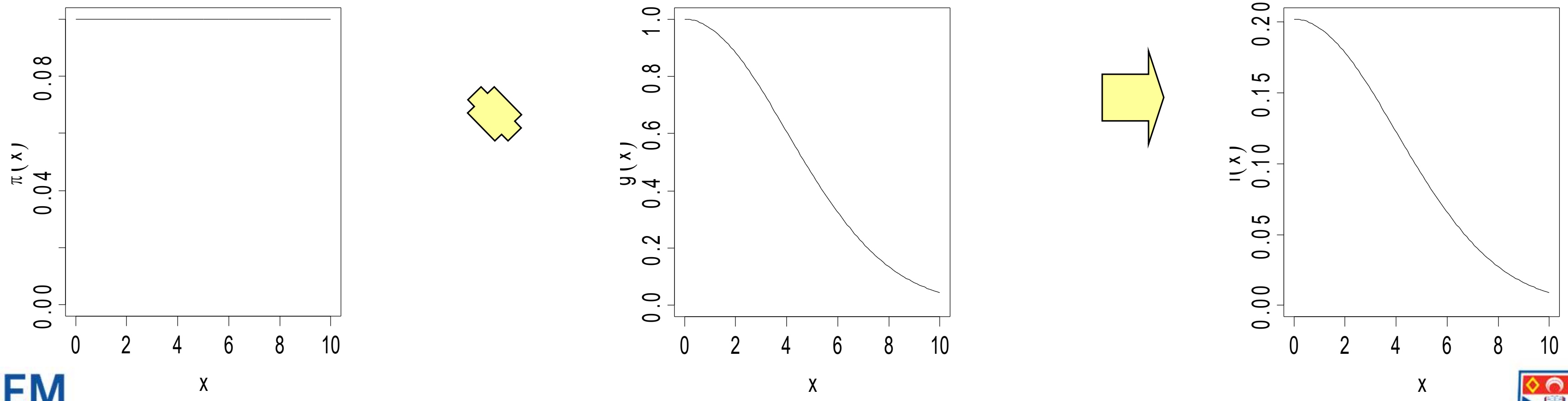
University of St Andrews

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

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

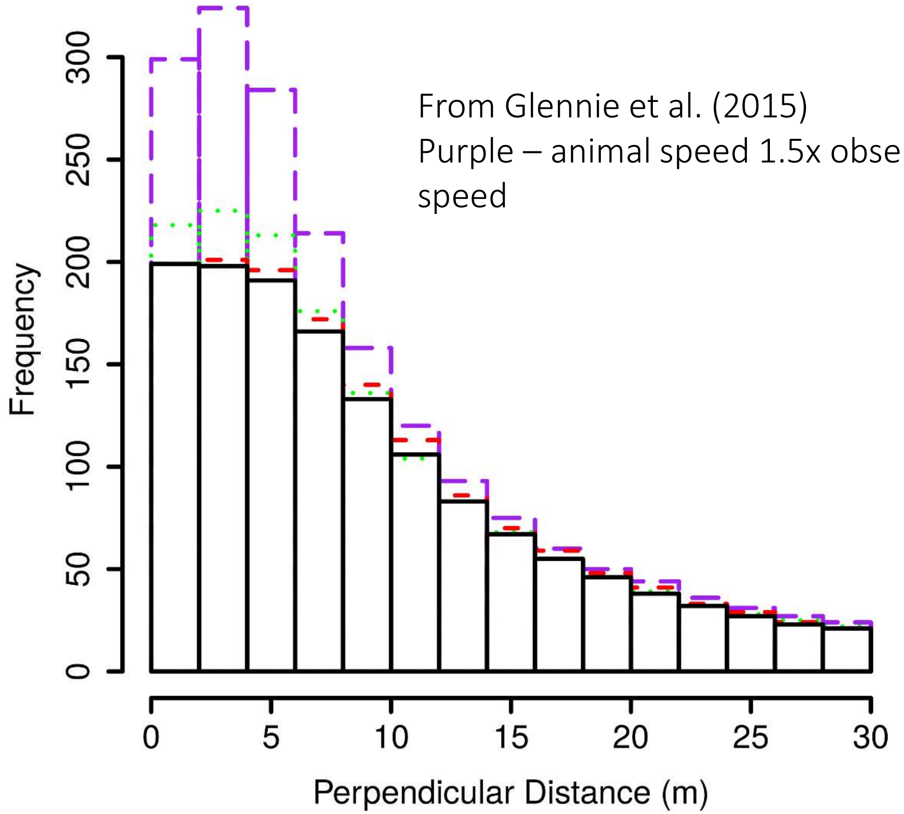
Animals do not move before they can be detected

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.

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



From Glennie et al. (2015)
Purple – animal speed 1.5x observer speed

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.

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 mildly affected)

Remedy to model selection challenge is addressed in

Howe, E. J., Buckland, S. T., Després-Einspenner, M.-L., & Kühl, H. S. (2019). Model selection with overdispersed distance sampling data. Methods in Ecology and Evolution, 10(1), 38–47. https://doi.org/10.1111/2041-210X.13082

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

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. Variance may also be biased.

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

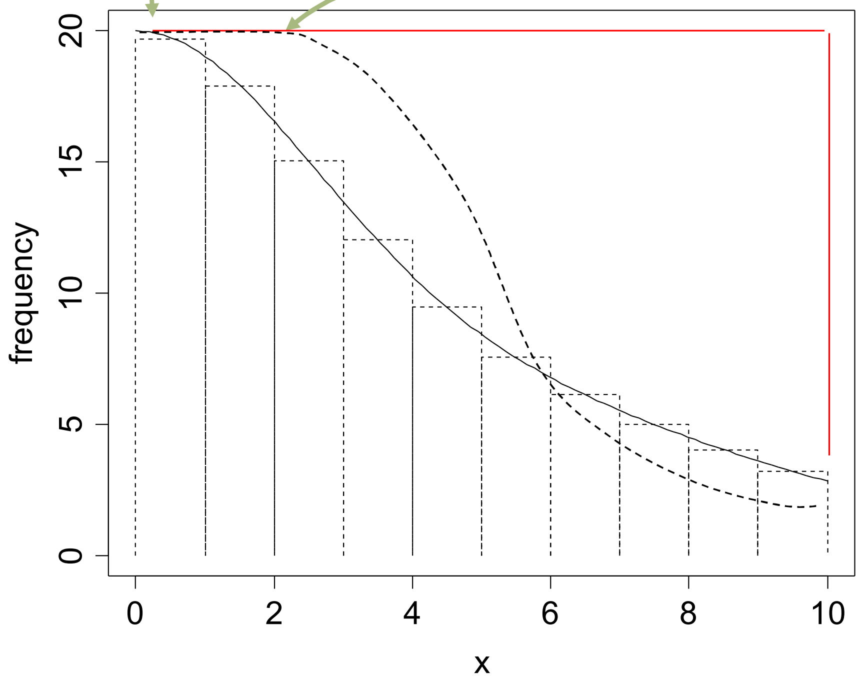
Reliable distance sampling (1)

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



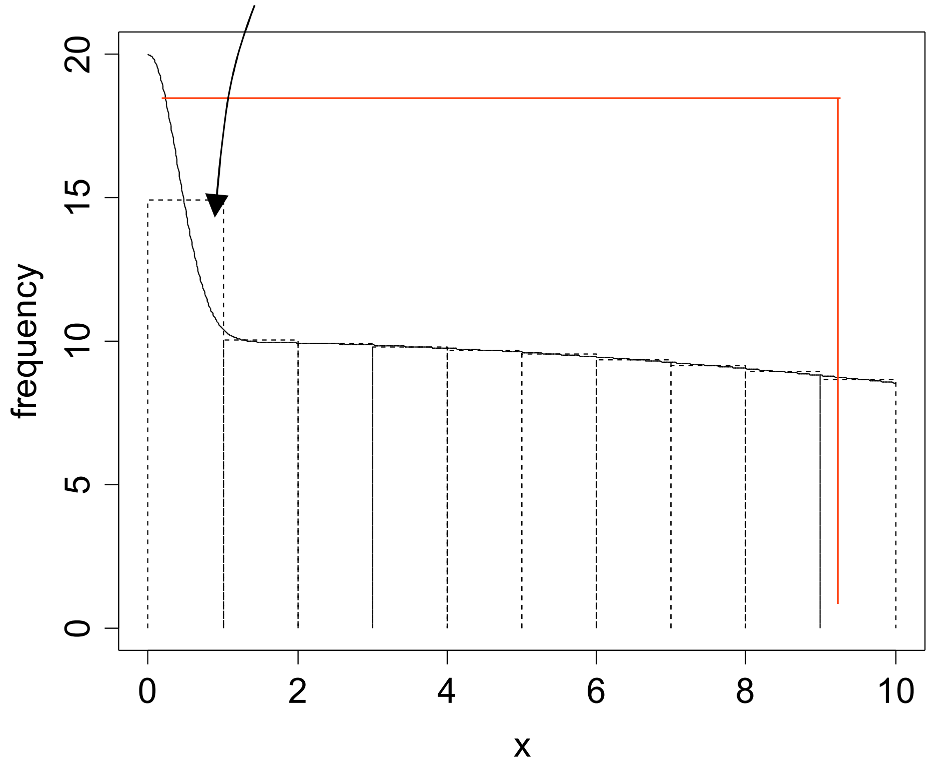
A wide shoulder makes it easier to estimate area under rectangle (or $f(0)$, etc.)

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

(1) Reliable estimation of P_a

Good field methods will avoid a 'spike' like this

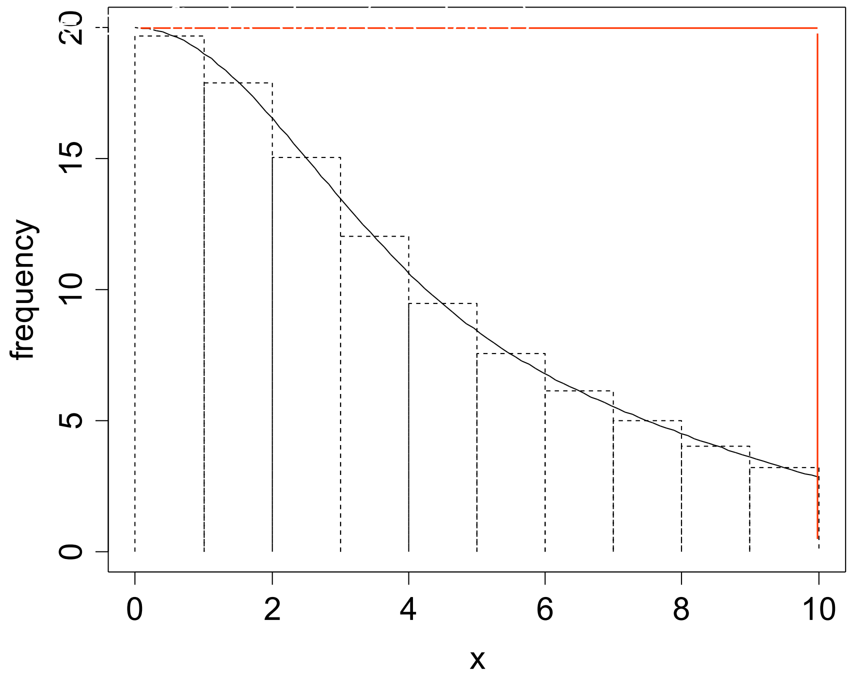


Avoid a) rounding distances (and angles) to zero,
b) 'guarding the trackline'

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

(1) Reliable estimation of P_a (cont.)



Sample size of observations (~60-80)
- less for detection functions with 'easy' shapes
- more for point transects and 'difficult shapes'.

CREEM Centre for Research into Ecological and Environmental Modelling

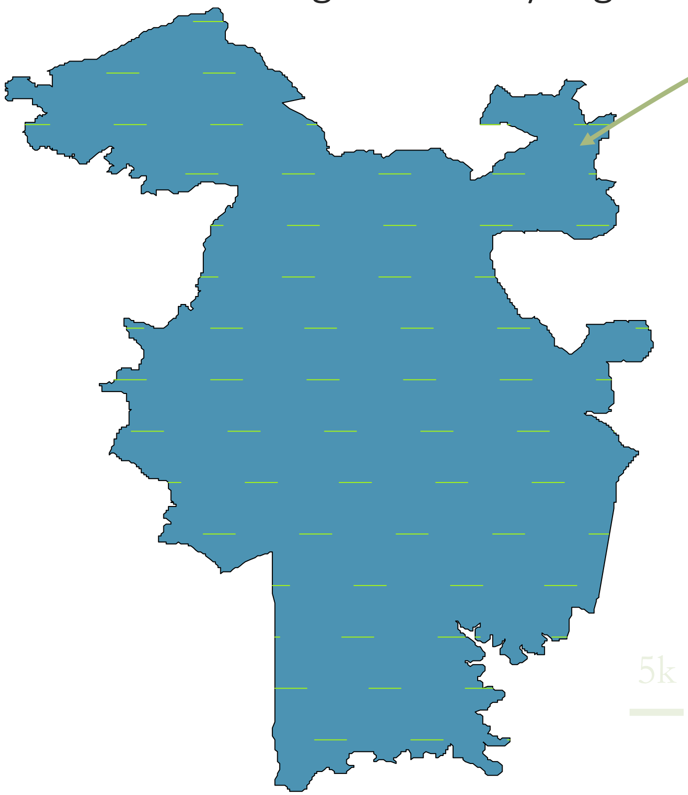
University of St Andrews

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



Photos: Ullas Karanth

CREEM Centre for Research into Ecological and Environmental Modelling

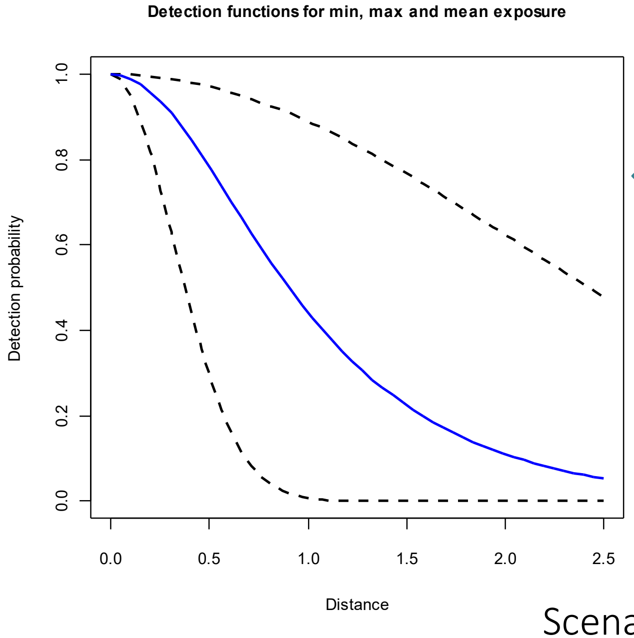
University of St Andrews

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

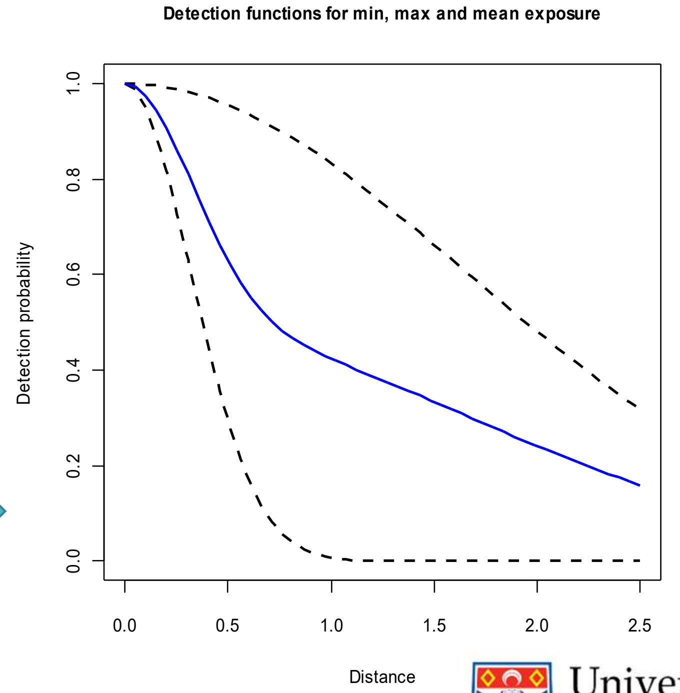


Scenario 1: animals have a gamma distribution of detection functions between min and max shown.

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