Assessment of model performance

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

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 $\log_e(L)$).

Akaike's Information Criterion

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

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.

- Models need not be special cases of one another
- 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 cutpoints 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 cutpoints for this 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.

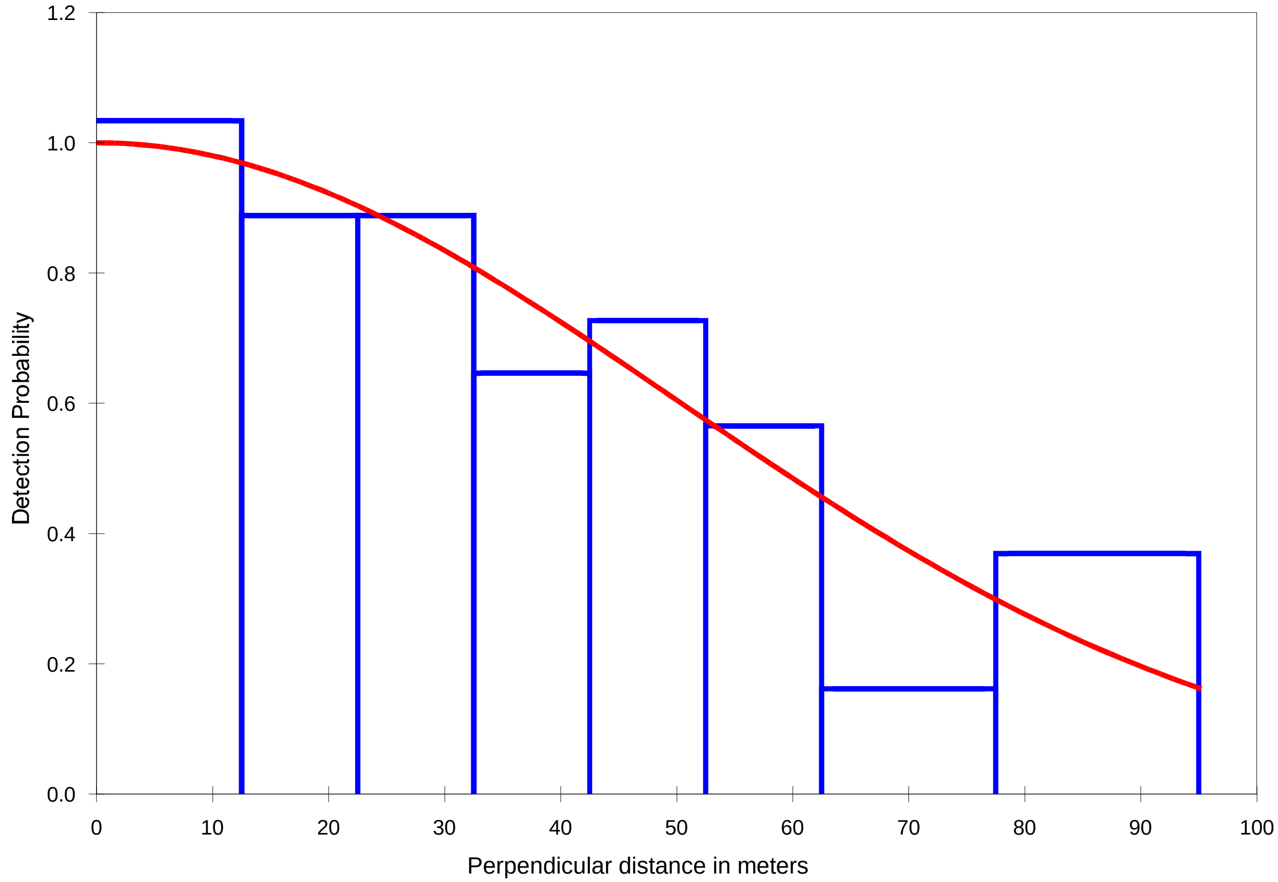
If the model is 'correct': $\chi^2 \sim \chi^2_{u-q-1}$

q = no. of parameters

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Chaffinch line transect data



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

χ^2 goodness-of-fit test

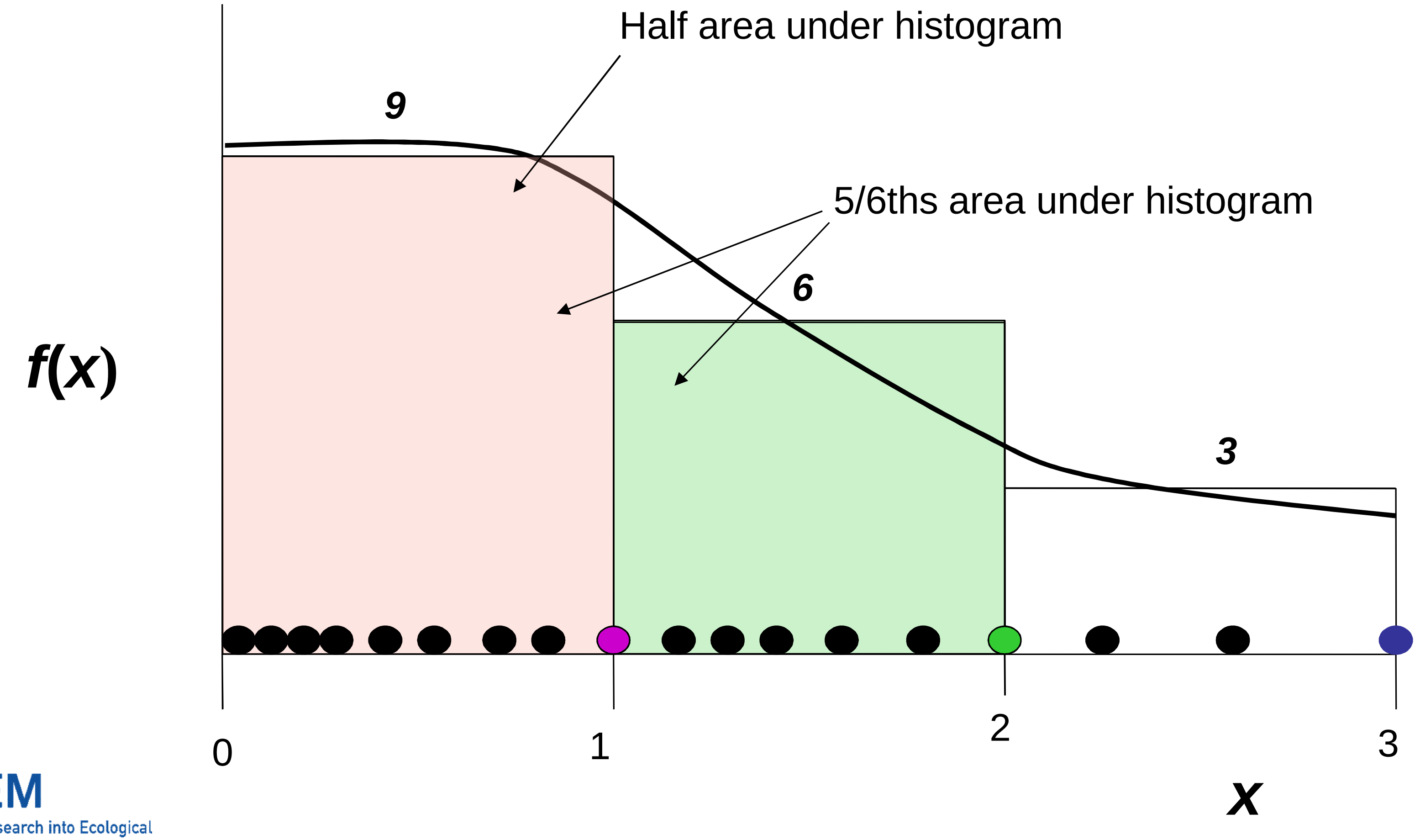
Cell i	Cut Points		Observed Values	Expected Values	Chi-square Values
1	0.000	12.5	16	15.32	0.030
2	12.5	22.5	11	11.63	0.034
3	22.5	32.5	11	10.62	0.013
4	32.5	42.5	8	9.33	0.189
5	42.5	52.5	9	7.87	0.164
6	52.5	62.5	7	6.37	0.062
7	62.5	77.5	3	6.96	2.253
8	77.5	95.0	8	4.91	1.953

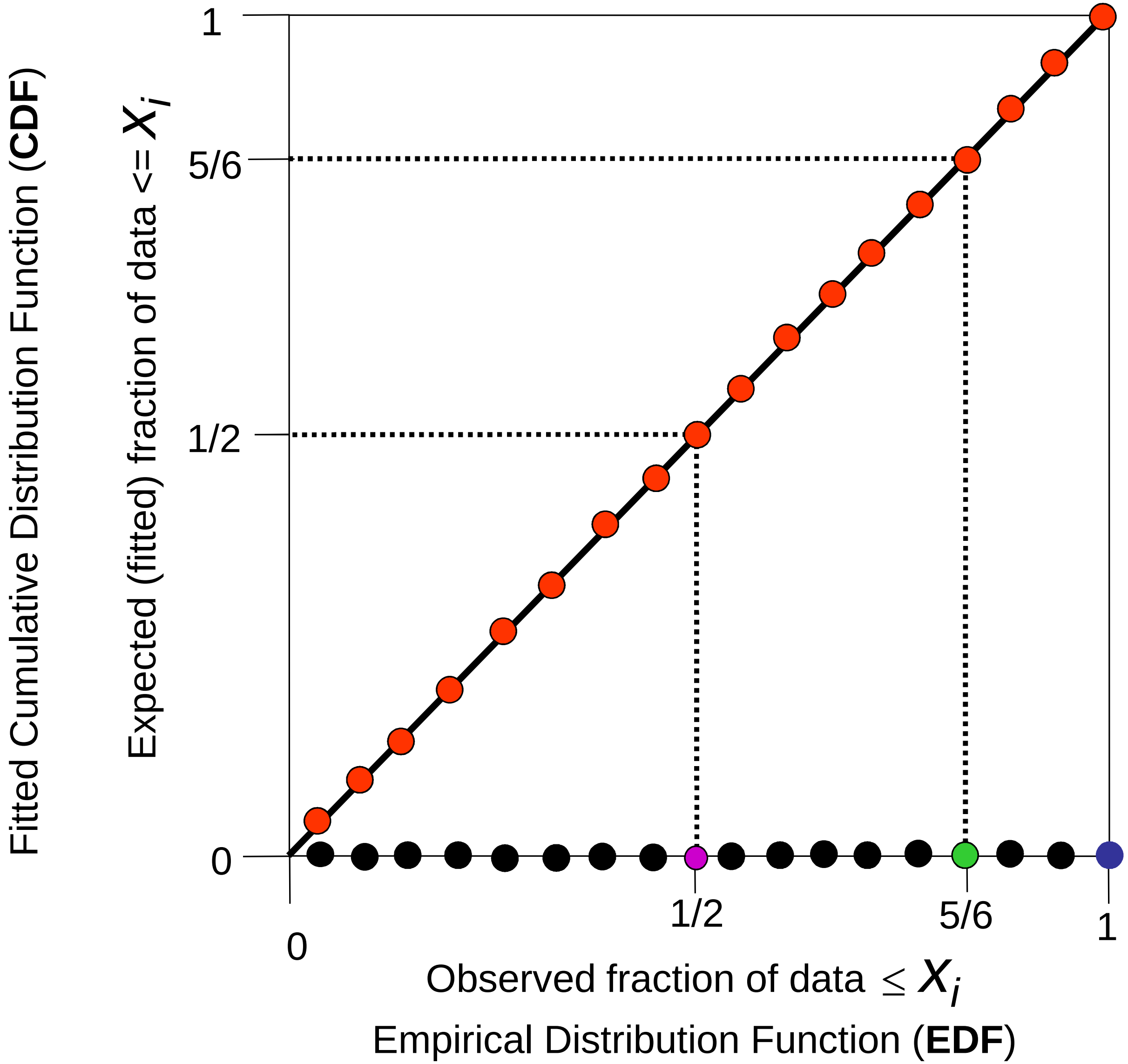
Total Chi-square value = 4.6970 Degrees of Freedom = 6.00

Probability of a greater chi-square value, P = 0.58322

The program has limited capability for pooling. The user should judge the necessity for pooling and if necessary, do pooling by hand.

Q-Q Plots and Related Tests

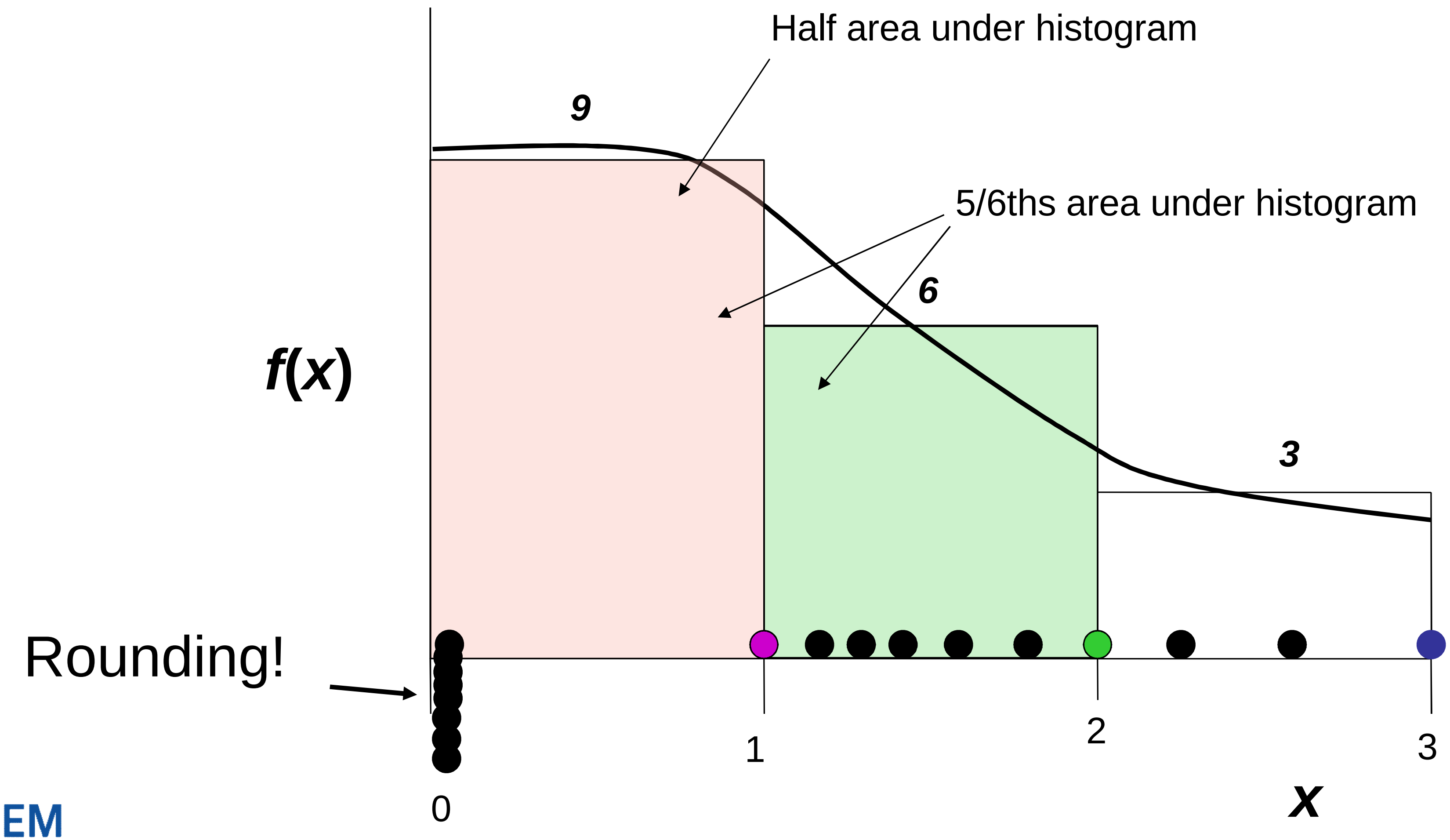




CREEM
Centre for Research into Ecological and Environmental Modelling

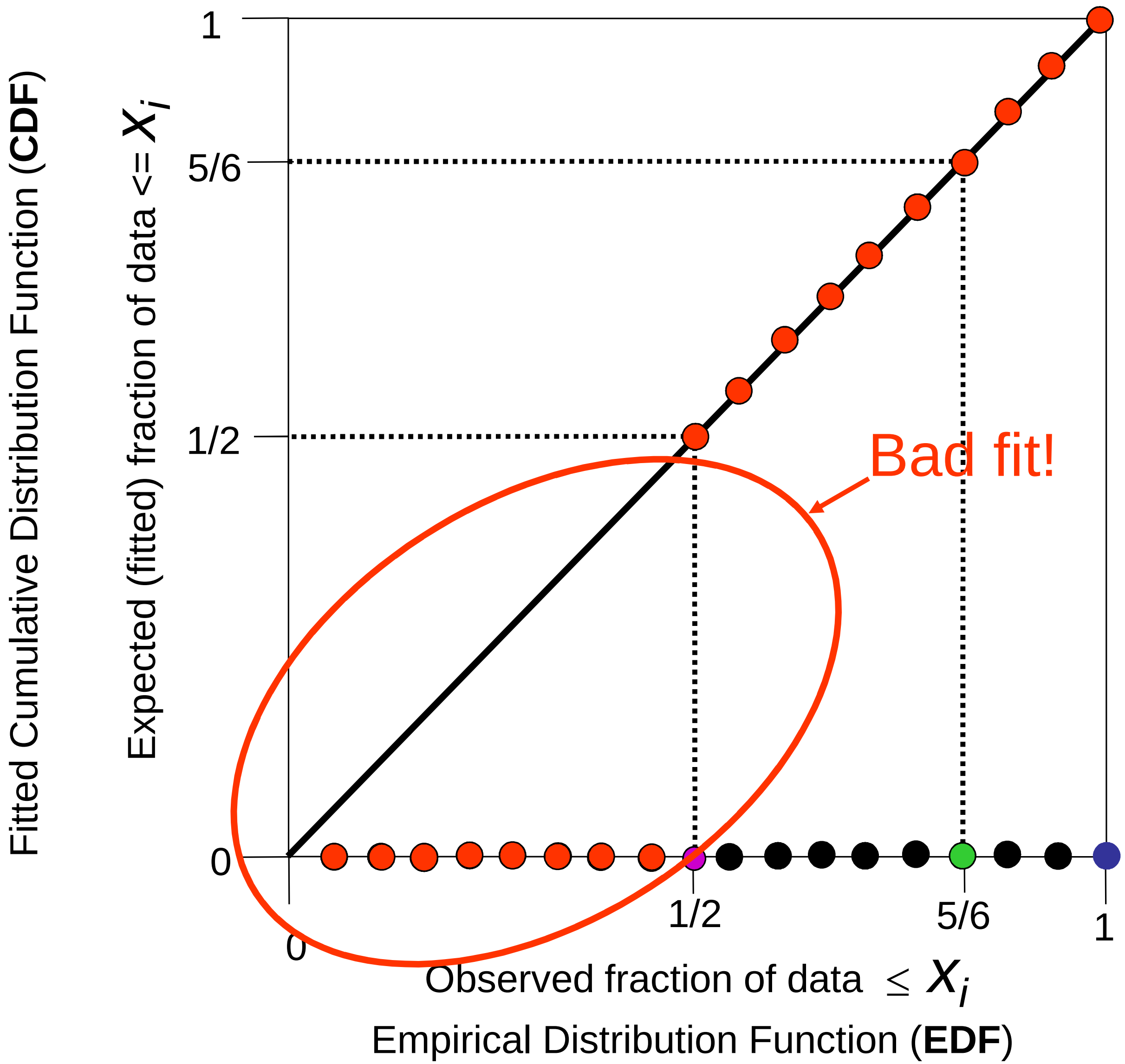
University of St Andrews

Example: Rounding to zero



CREEM
Centre for Research into Ecological and Environmental Modelling

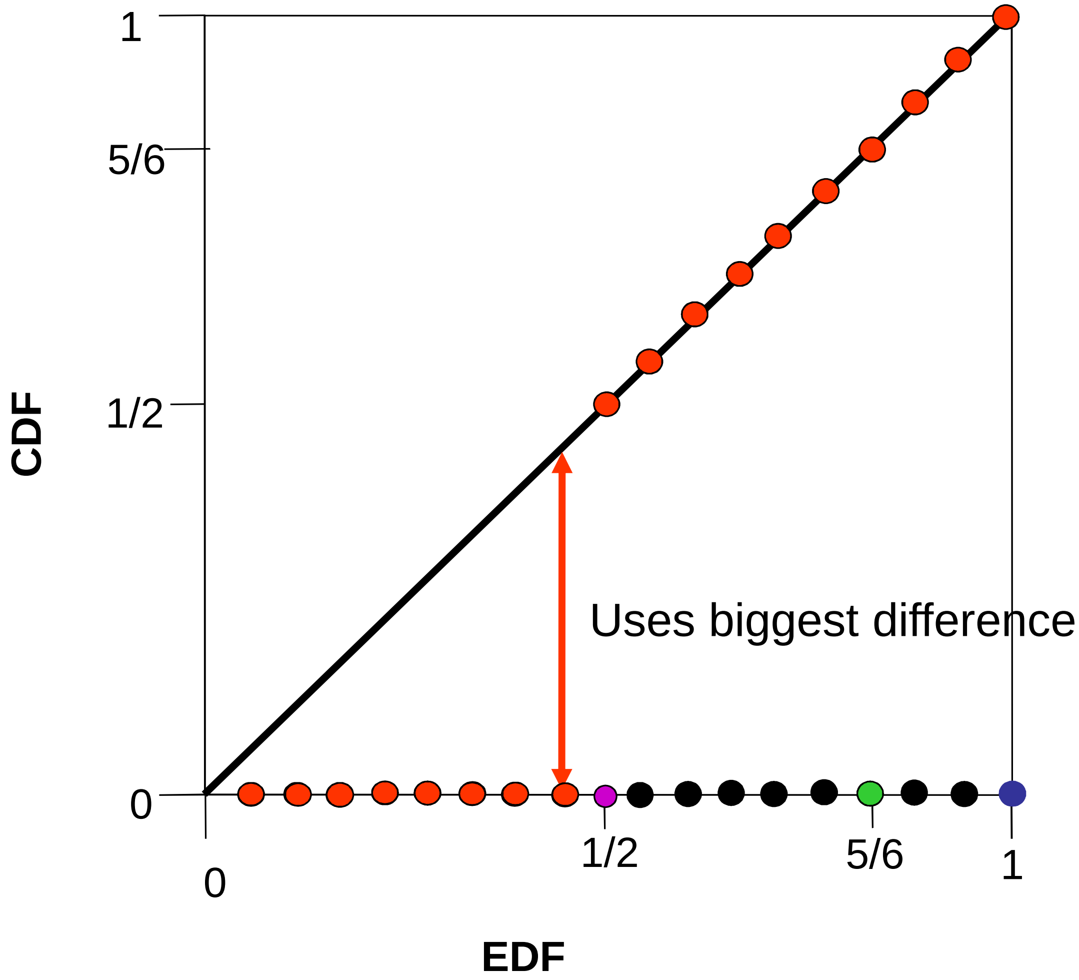
University of St Andrews



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

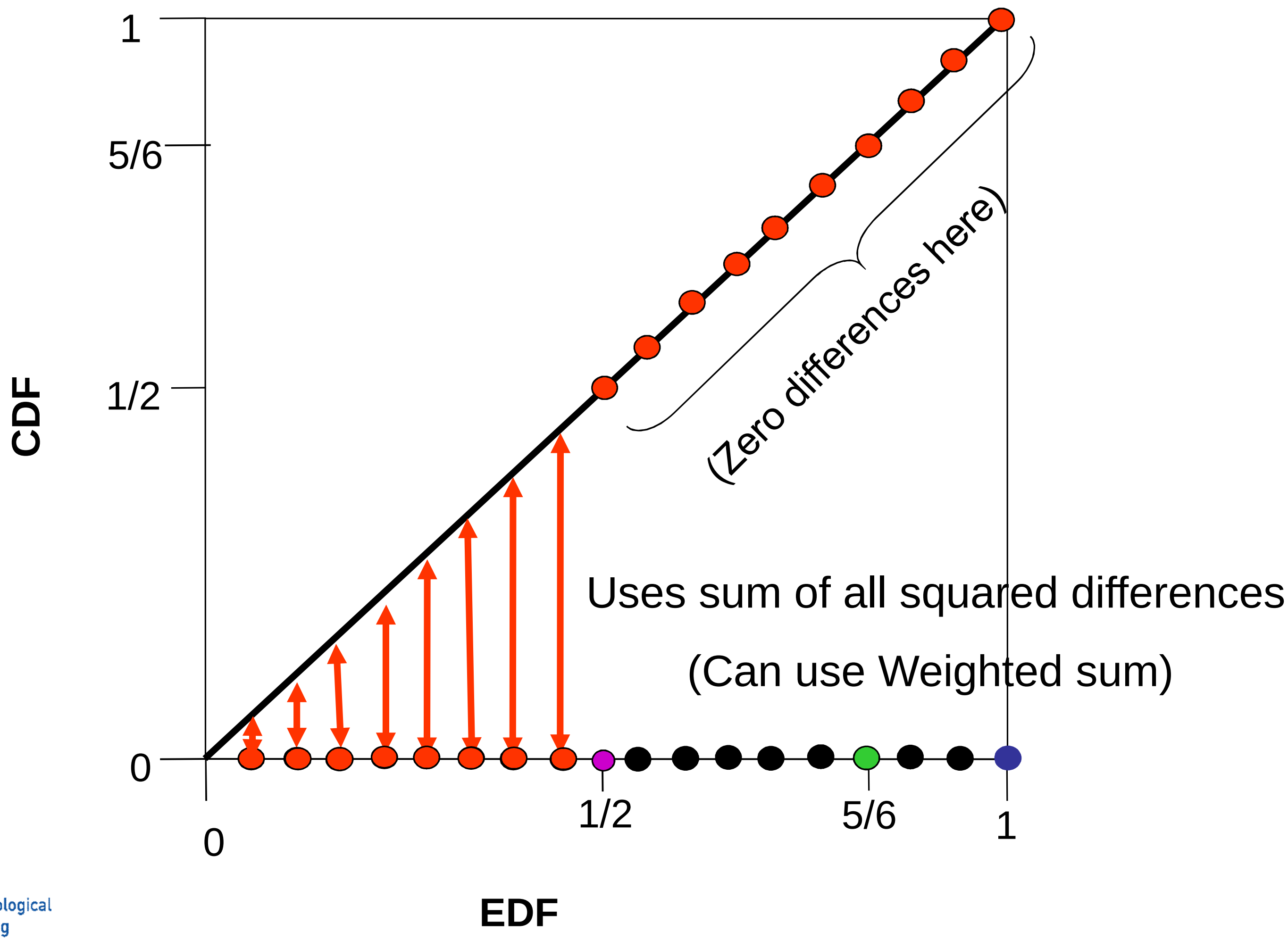
Kolmogorov-Smirnov test



CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

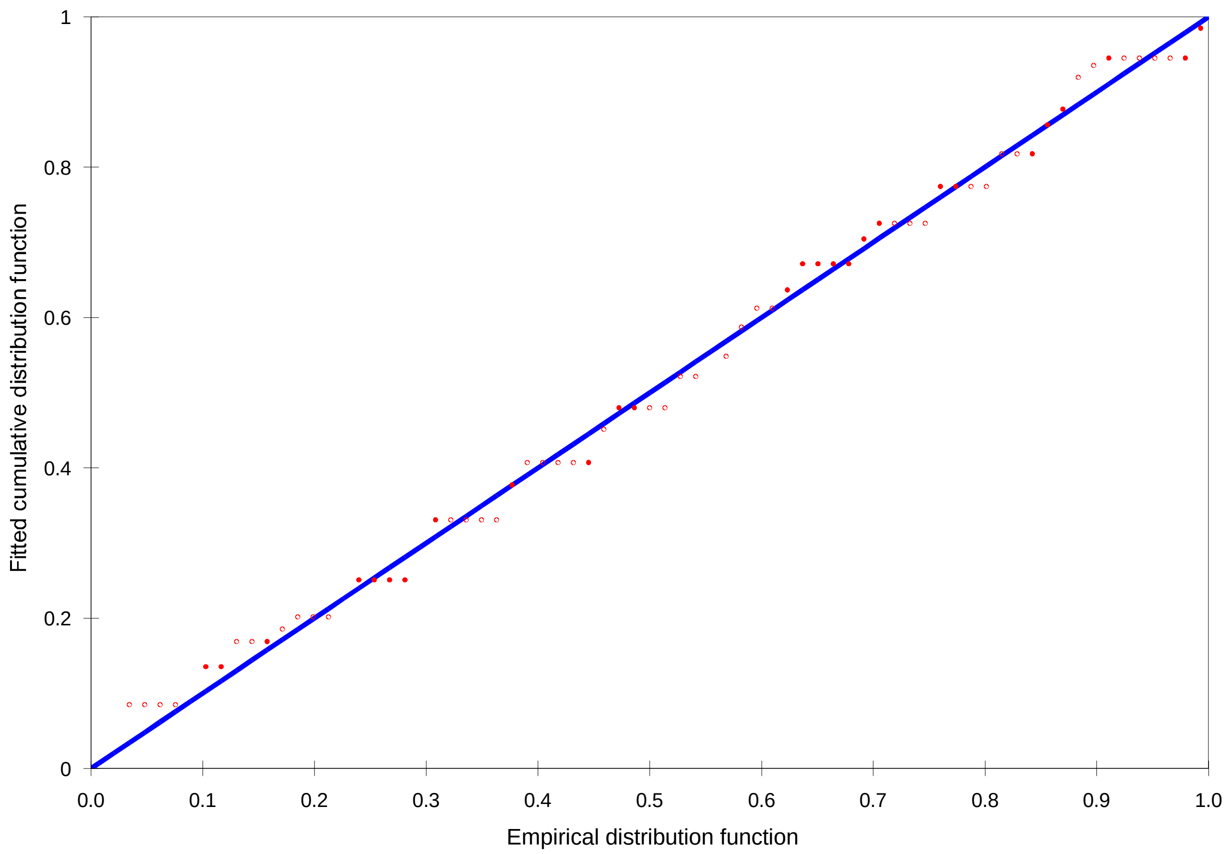
Cramér-von Mises test



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

```
Kolmogorov-Smirnov test
 -----------------------

 D_n                        = 0.0573                   p  = 0.9703


 Cramer-von Mises family tests
 -----------------------------

 W-sq (uniform weighting) = 0.0368          0.900 < p <= 1.000
   Relevant critical values:
     W-sq crit(alpha=0.900) = 0.0000

 C-sq (cosine weighting)  = 0.0257          0.900 < p <= 1.000
   Relevant critical values:
     C-sq crit(alpha=0.900) = 0.0000
```

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
- Cramér-von Mises test can be weighted, to give higher weight to *x* near zero