

# Choosing a Detection function

# Overview

Formal definition

Criteria for a good detection function model

Key functions and adjustment terms

Fitting models in Distance

Choosing the number of parameters

Introduction to truncation

## Formal definition

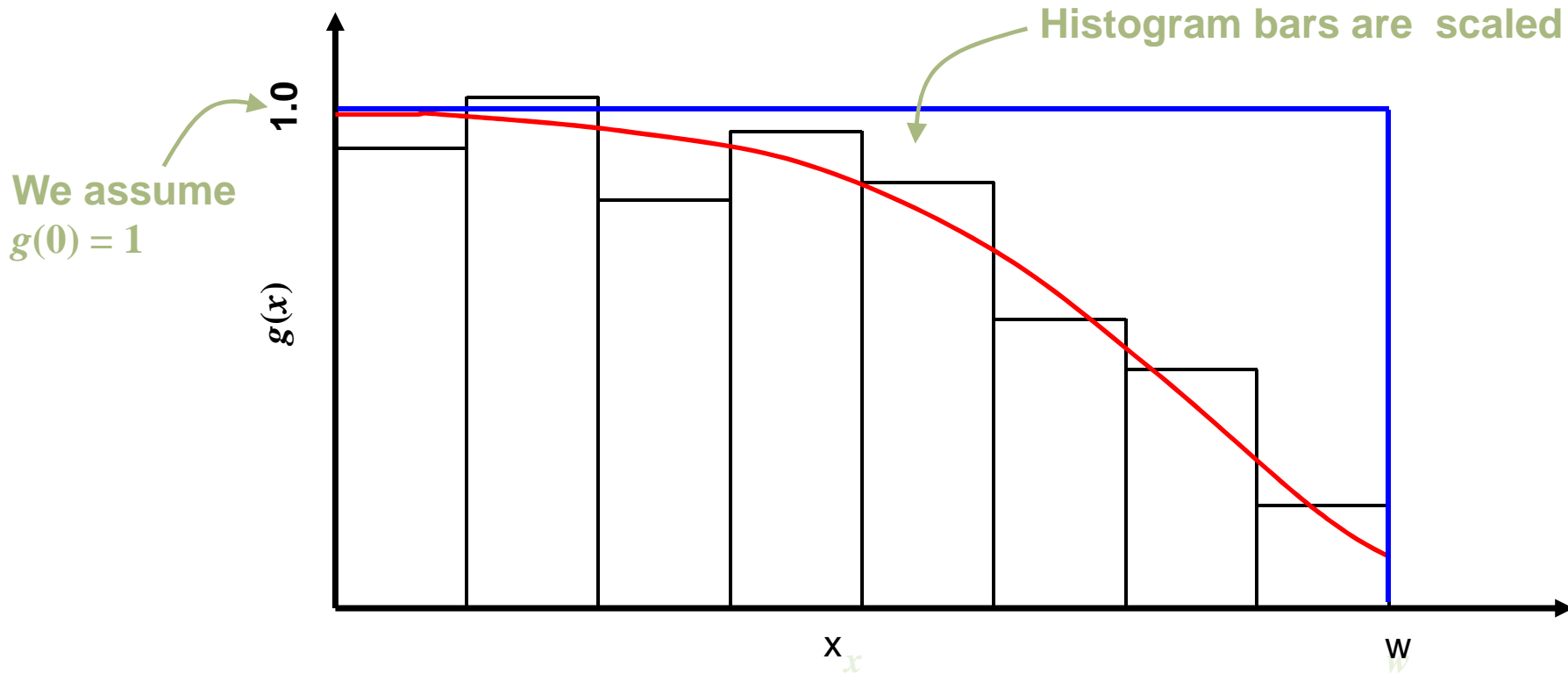
The **detection function** describes the relationship between distance and the probability of detection

Formally denoted by  $g(x)$  (usually referred to as 'g of x')

**$g(x)$  = the probability of detecting an animal, given that it is at distance  $x$  from the line**

Key to the concept of distance sampling

# The detection function, $g(x)$



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

## Modelling $g(x)$

$g(x)$  represents the **underlying** relationship between detection probability and distance

However, the true form of  $g(x)$  is unknown to us

We need to **estimate**  $g(x)$  by fitting a **model** to our data

i.e., we need to find a curve that will approximate the underlying relationship

# Criteria for robust estimation

Four main criteria for a good model:

1. **Model robustness** – use a model that will fit a wide variety of plausible shapes for  $g(x)$
2. **Shape criterion** – use a model with a ‘shoulder’ – i.e.  $g'(0)=0$
3. **Pooling robustness** – use a model for the average detection function, even when many factors affect detectability
4. **Estimator efficiency** – use a model that will lead to a precise estimator of density

# Key functions

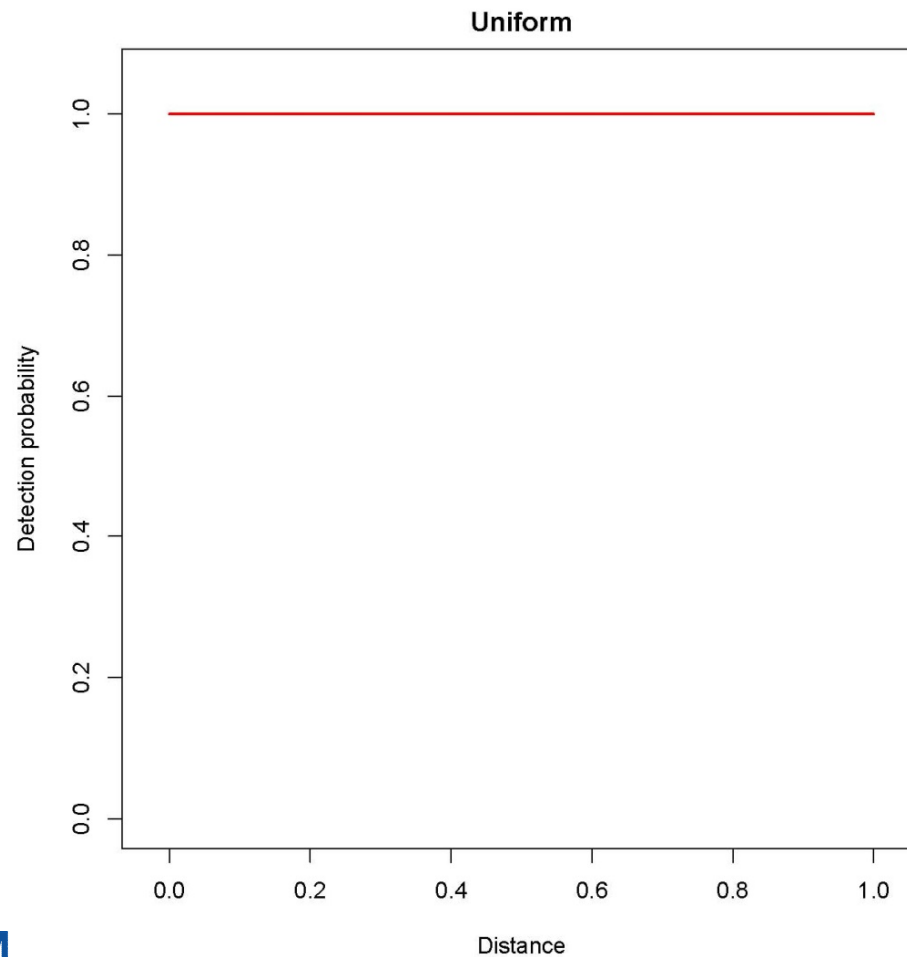
The first step in constructing a model for  $g(x)$  is to choose a **key function**

This determines the basic model shape

Four key functions available in Distance:

1. Uniform
2. Half normal
3. Hazard rate
4. Negative exponential

# Key functions (cont.)



- Model formula:

$$g(x) = 1, x \leq w$$

- Parameters = 0
- Shape criterion?

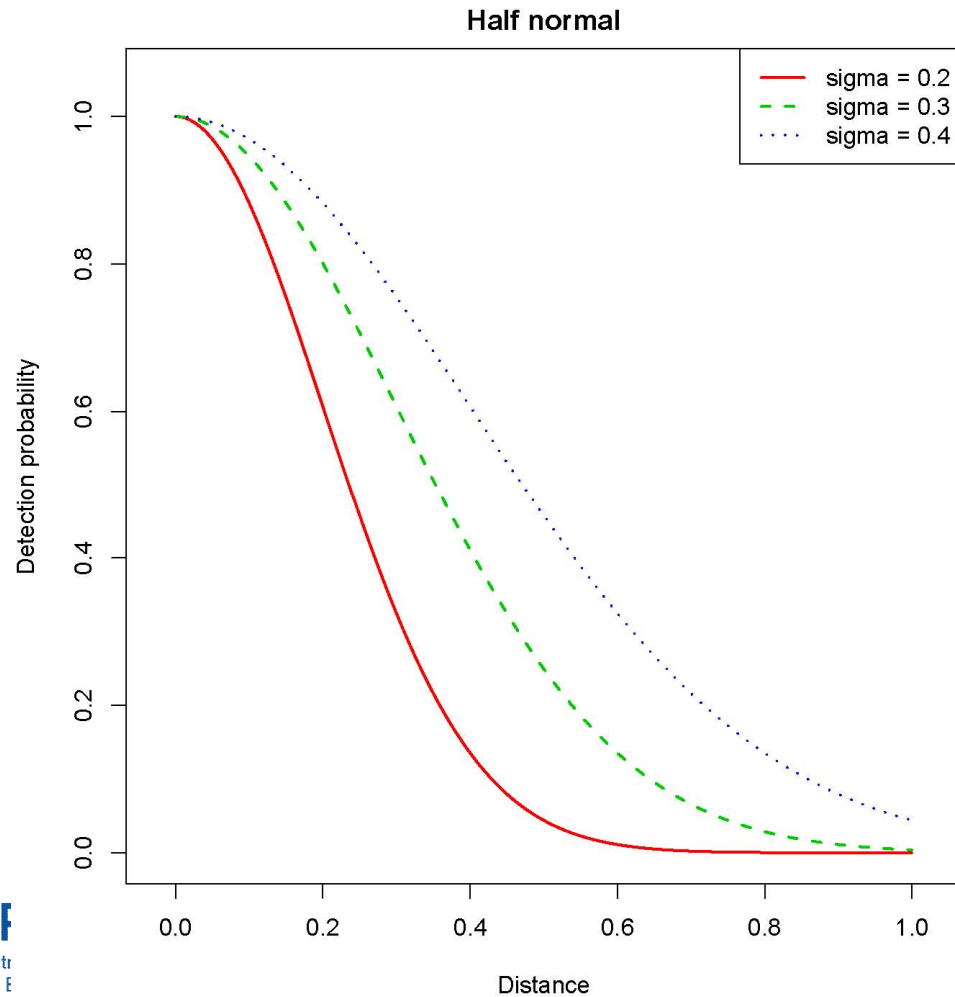
Yes

- Model robust?

No



## Key functions (cont.)



- Model formula:

$$g(x) = \exp\left(\frac{-x^2}{2\sigma^2}\right), x \leq w$$

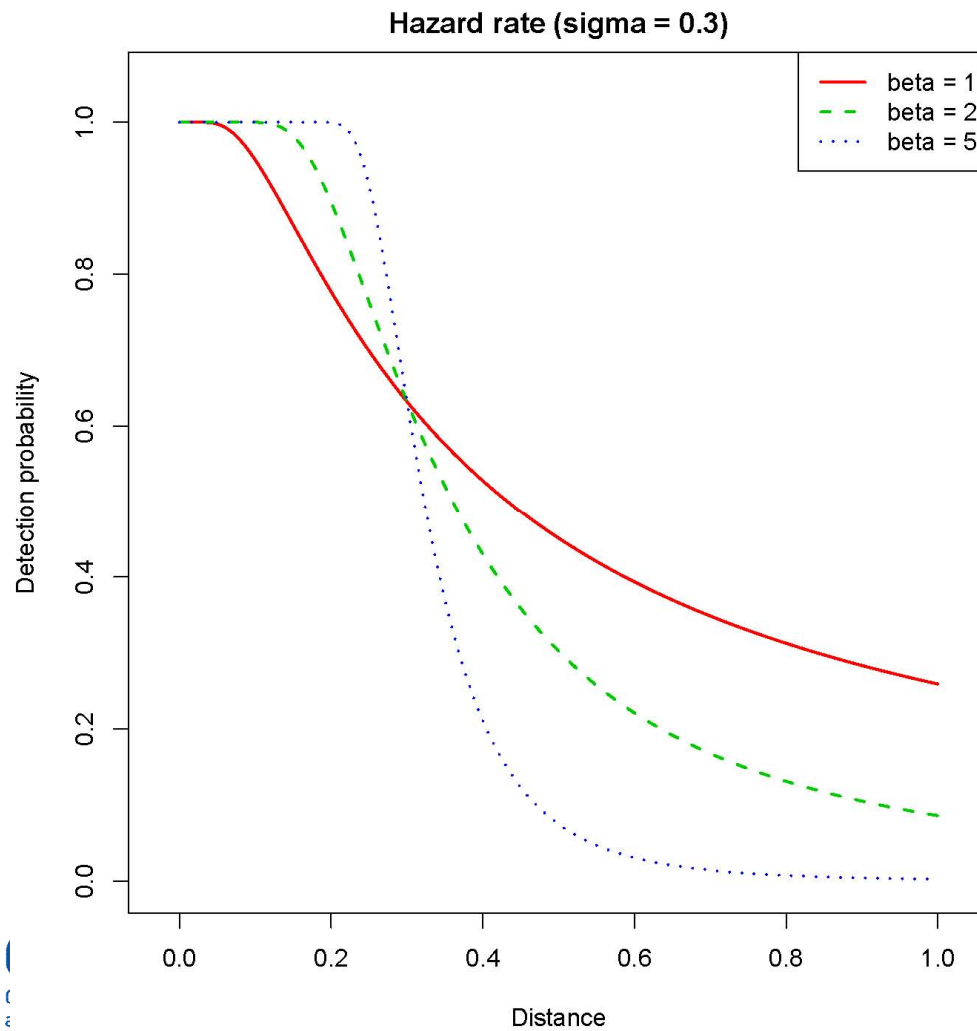
- Parameters = 1
- Shape criterion?

Yes

- Model robust?

No

## Key functions (cont.)



- Model formula:
$$g(x) = 1 - \exp\left[-\left(\frac{x}{\sigma}\right)^{-\beta}\right], x \leq w$$

- Parameters = 2
- Shape criterion?

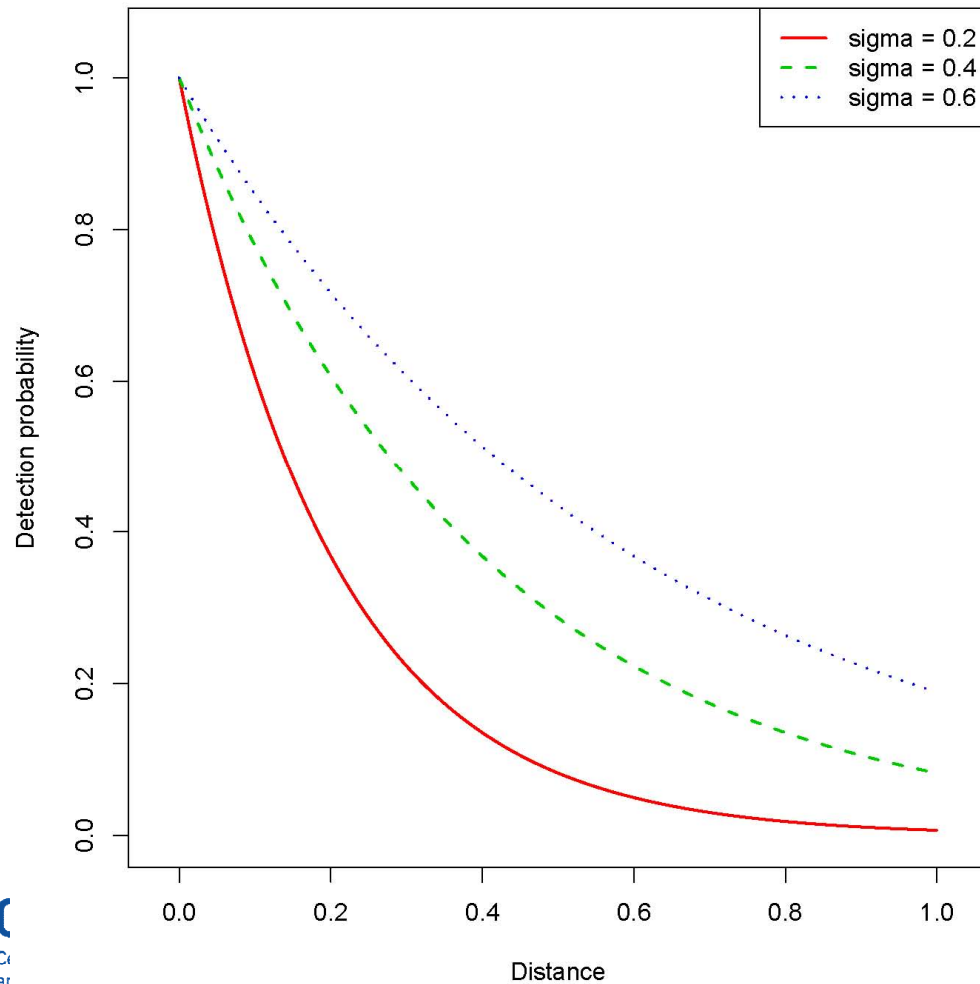
Yes

- Model robust?

Yes

# Key functions (cont.)

Negative exponential



- Model formula:

$$g(x) = \exp\left(\frac{-x}{\sigma}\right), x \leq w$$

- Parameters = 1
- Shape criterion?

**No**

- Model robust?

**No**

# Key functions in Distance

The screenshot displays the Distance software interface. On the left, the 'Analysis 139' window shows the 'Model definition' list with '5 HN by strat f0 pooled' selected. The 'Properties ...' button next to it is circled in red. An arrow points from this button to the 'Model Definition Properties' dialog box on the right. In the dialog, the 'Key function' dropdown is set to 'Half-normal' and is also circled in red. The 'Series expansion' is set to 'Cosine'. The 'Analysis Engine' is 'CDS - Conventional distance sampling'. The 'Models' tab is active, and the 'Detection function models' table is shown below.

Model	Key function	Series expansion	
1	Half-normal	Cosine	-

Select among multiple models using: AIC

# Adjustment terms

Models can be made more robust by adding a series of **adjustment terms** (also called **series expansion** or **series adjustment**) to the key function

Key function  $\times$  (1 + Series)

Series =  $\alpha_1 \times \text{term}_1 + \alpha_2 \times \text{term}_2 + \dots$  etc.

The  $\alpha_i$  parameters must be estimated

Resulting curve model is scaled so that  $g(0)=1$

The number of adjustment terms needs to be chosen

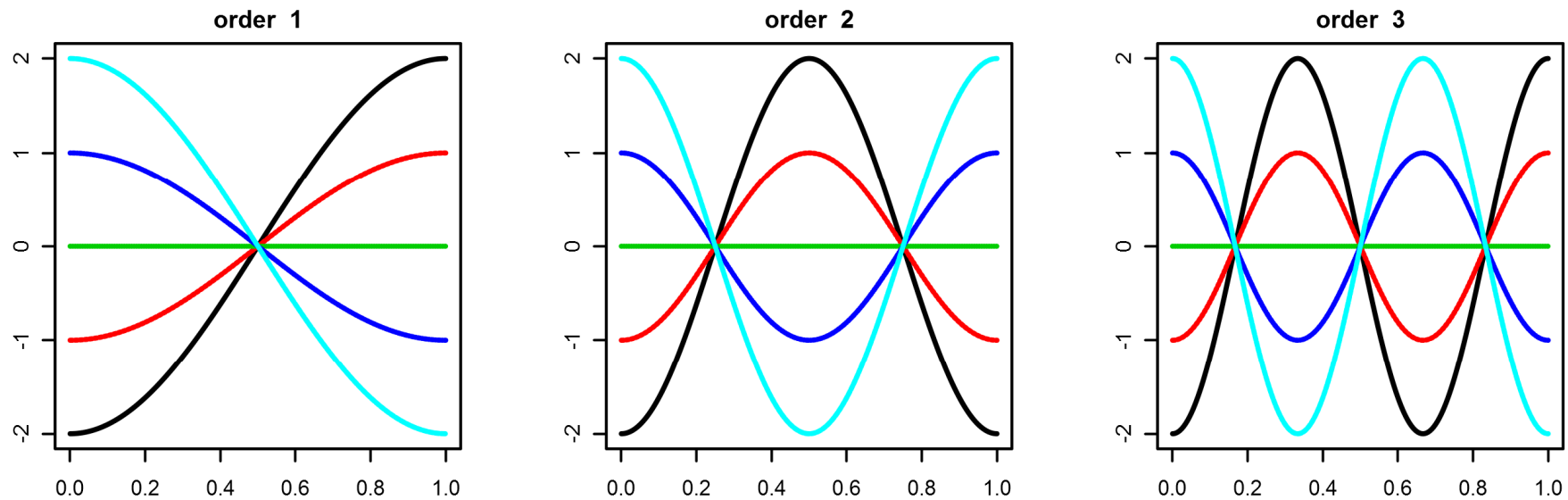
# Adjustment terms

Distance allows the selection of three types of series (one type per model)

Key function	Series adjustment
Uniform*	Cosine*
Half normal <sup>†</sup>	Hermite polynomial <sup>†</sup>
Hazard rate	Simple polynomial
Negative exponential	

# How adjustment terms work

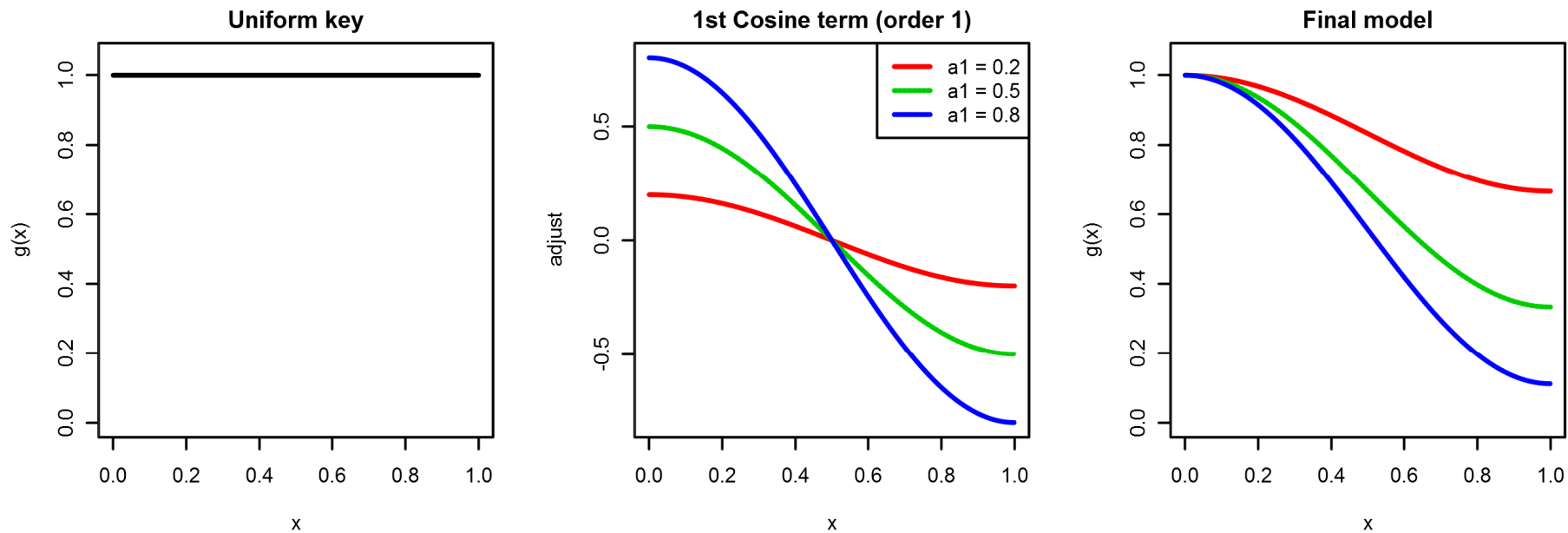
E.g. Cosine series (for different values of  $\alpha$ )



(1<sup>st</sup> order only used for uniform)

# How adjustment terms work

E.g. Uniform + 1 Cosine adjustment term:

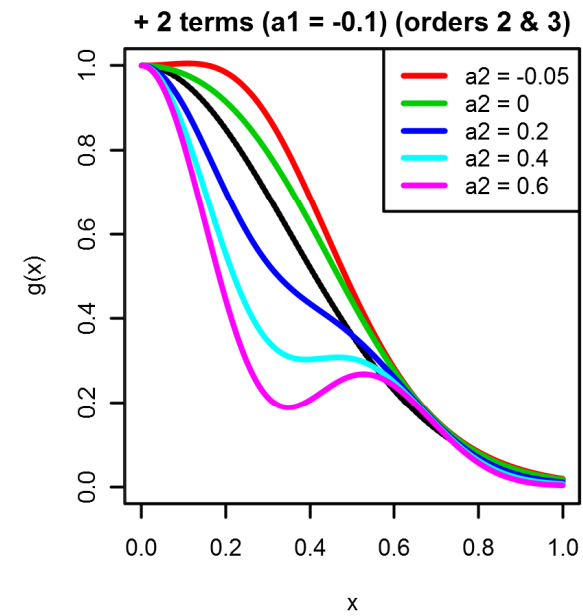
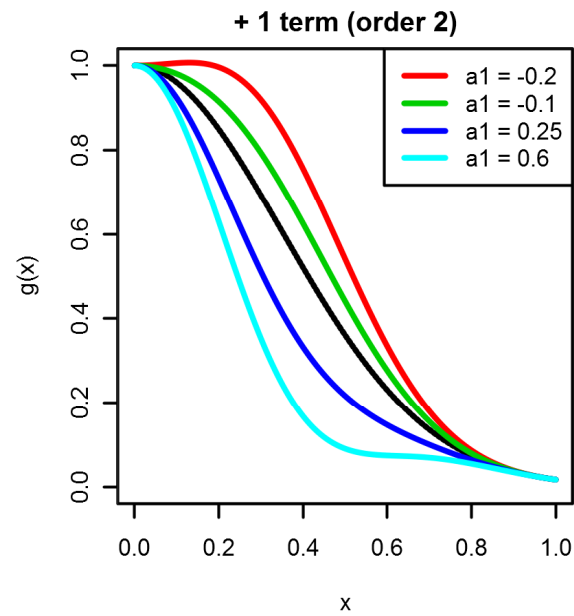
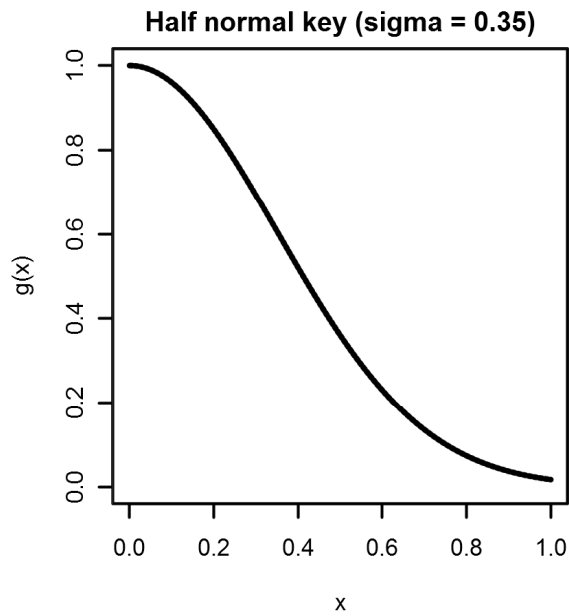


The effect of the adjustment terms depends on the value of their parameters



# How adjustment terms work

E.g. Half normal + 1 or 2 Cosine terms:



# Adjustments in Distance

The screenshot displays the 'Model Definition Properties: [HN by strat f0 pooled]' dialog box in the Distance software. The 'Analysis Engine' is set to 'CDS - Conventional distance sampling'. The 'Models' tab is selected, showing a table of detection function models. The first model is highlighted, with 'Half-normal' as the key function and 'Cosine' as the series expansion. The 'Properties ...' button in the main window is also circled, with an arrow pointing to the 'Models' tab in the dialog box.

Analysis: Name: a1 - HN by strat f0 pooled w82.5  
Created: 20-12-2005 10:38:48  
Run: 25-01-2008 15:47:48

Survey: Set 1 [1] New Survey

Data filter:  
1 Default Data Filter  
2 Trunc 82.5

Model definition:  
1 HRpol by strat  
2 HN MCDS Obs by strat  
3 HN by strat  
4 HR MCDS Obs by strat  
5 HN by strat f0 pooled  
6 HR by strat f0 pooled  
7 Uni+cos by strat  
8 Unicos by strat f0 pooled

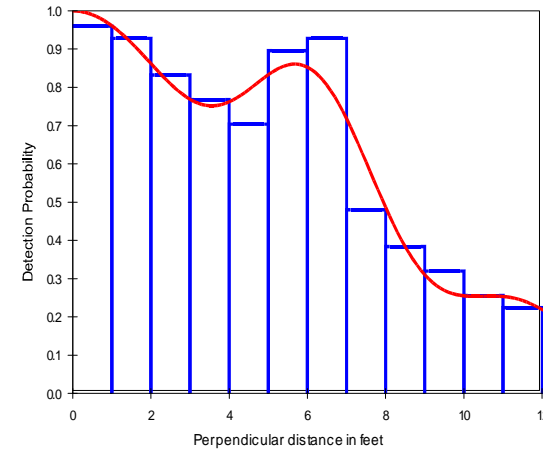
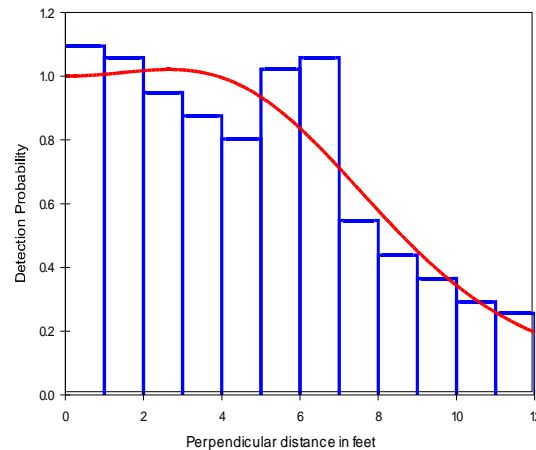
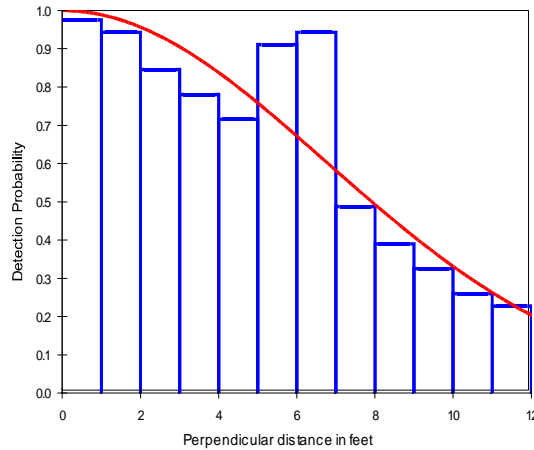
Model Definition Properties: [HN by strat f0 pooled]  
Analysis Engine: CDS - Conventional distance sampling  
Estimate Detection function Cluster size Multipliers Variance Misc.  
Models Adjustment terms Constraints Diagnostics  
Detection function models:  
Model Key function Series expansion  
1 Half-normal Cosine  
Select among multiple models using AIC

# Adjustments in Distance

The image shows a software interface for model definition. On the left, the 'Analysis' window displays details for 'a1 - HN by strat f0 pooled w82.5'. The 'Model definition' list includes '5 HN by strat f0 pooled', which is highlighted. A red circle around the 'Properties ...' button for this model points to the 'Model Definition Properties' dialog box on the right.

The 'Model Definition Properties' dialog box is titled 'Model Definition Properties: [HN by strat f0 pooled]'. It shows the 'Analysis Engine' set to 'CDS - Conventional distance sampling'. The 'Adjustment terms' tab is selected and highlighted with a red box. Under 'Selection of adjustment terms', 'Automated selection' is chosen. The 'Selection method' is 'Sequential', 'Look-ahead' is '1', 'Selection criterion' is 'AIC', 'Significance level' is '0,15', and 'Maximum terms' is '5'. There are two empty tables for 'Manual selection' and 'Manually select starting values'. The 'Scaling of distances' section shows 'Scale distances by:' set to 'W (right truncation distance)'. The 'Name' field at the bottom is 'HN by strat f0 pooled'.

# Adjustment terms – how many?



Half normal	Half normal	Half normal
0 adjustment terms	1 adjustment term	5 adjustment terms
1 parameter	2 parameters	6 parameters
$\hat{P}_a = 0.65$	$\hat{P}_a = 0.72$	$\hat{P}_a = 0.63$
$CV(\hat{P}_a) = 5.8\%$	$CV(\hat{P}_a) = 11.6\%$	$CV(\hat{P}_a) = 19.9\%$

**Note:** There is a monotonicity constraint in Distance that is switched on by default to prevent detection functions from increasing. The constraint had to be turned off to produce the third plot. The third plot is for demonstration only – it would not be a good detection function to choose (unless there was a biological reason why detection probability would increase at those distances).

# How many parameters?

Models with too few parameters will not be flexible enough to describe the underlying relationship

Adding parameters will improve the fit

But models with too many parameters will be too flexible and will also describe the random noise in the data

We generally require models with an intermediate number of parameters

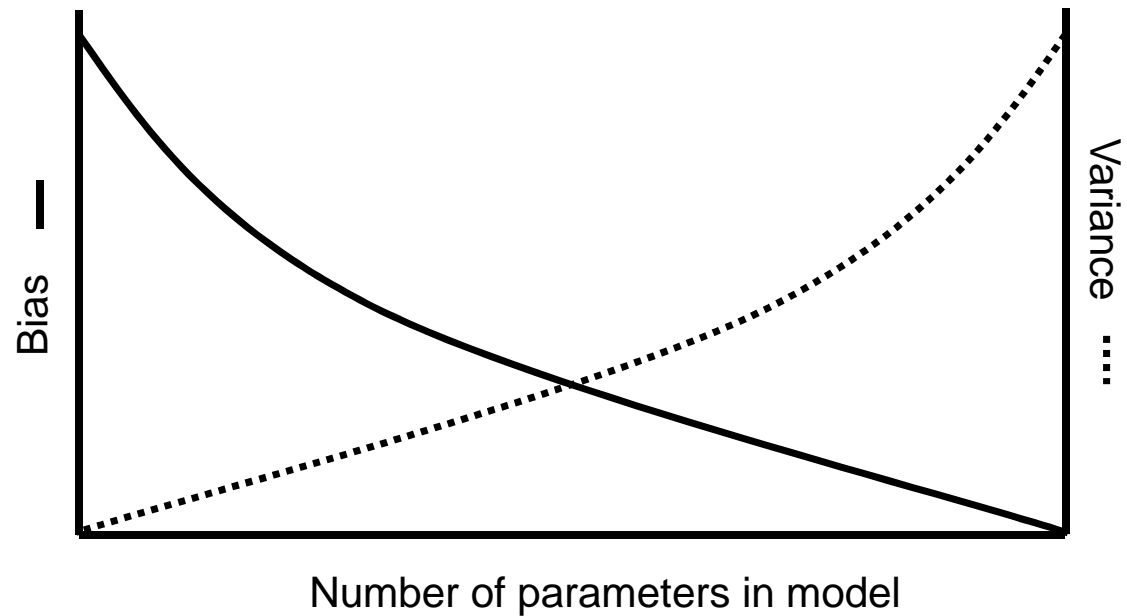
## How many parameters?

This problem can also be expressed as a trade-off between bias and variance

Models with too few parameters tend to produce estimates with low variance and high bias

Models with too many parameters tend to produce estimates with low bias and high variance (note the increasing CV for the estimate of  $P_a$  on the previous slide)

# How many parameters?



Need an objective way of choosing the 'best' model...

# Truncation

$$\hat{N} = \frac{nA}{2wL\hat{P}_a}$$

Need to choose the value of  $w$  (right truncation)

Large distances contribute little to estimating the shape of  $g(x)$  at small distances (i.e. the shoulder) and may lead to poor fit and high variance

Typically we might truncate around 5% of observation for line transects (perhaps nearer 10% for point transects)

Can truncate in the field or at the analysis stage