

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

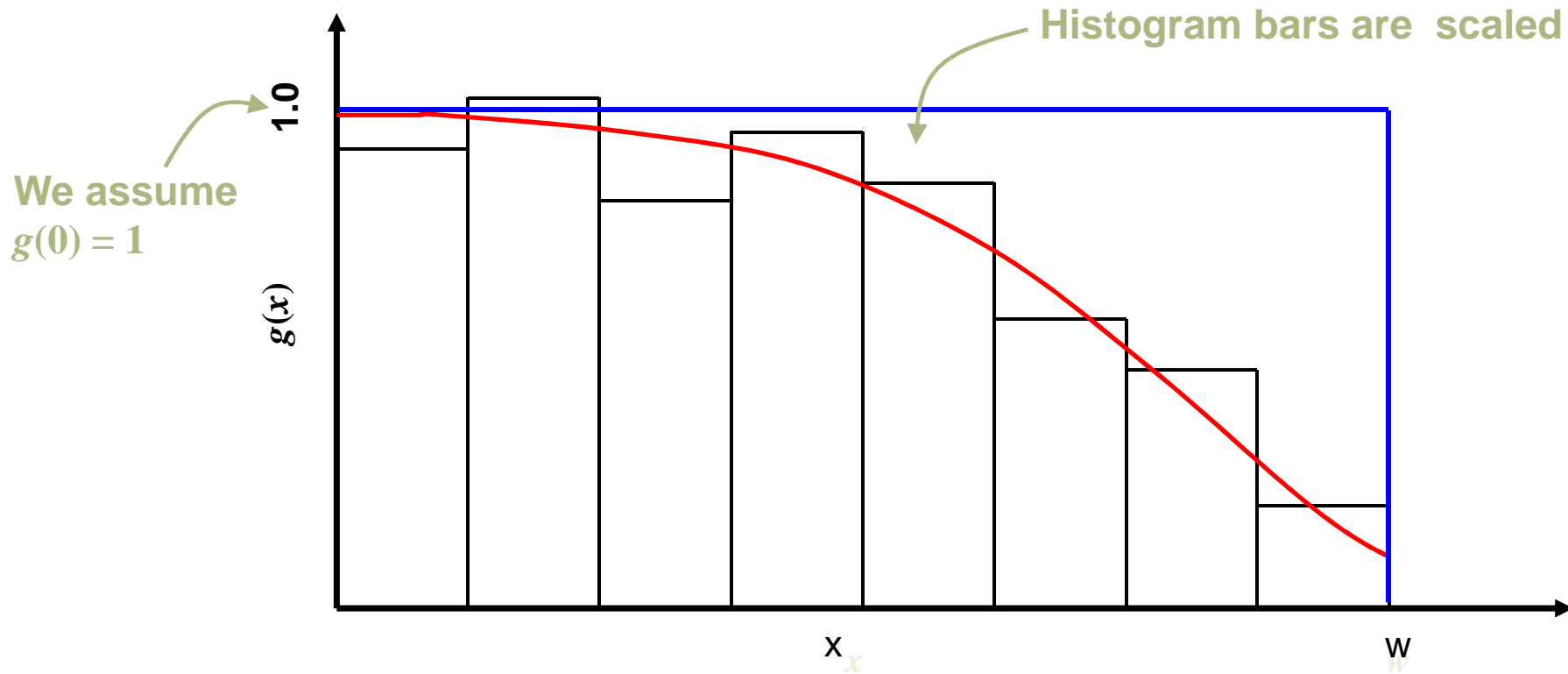
The **detection function** describes the relationship between distance and the probability of detection

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

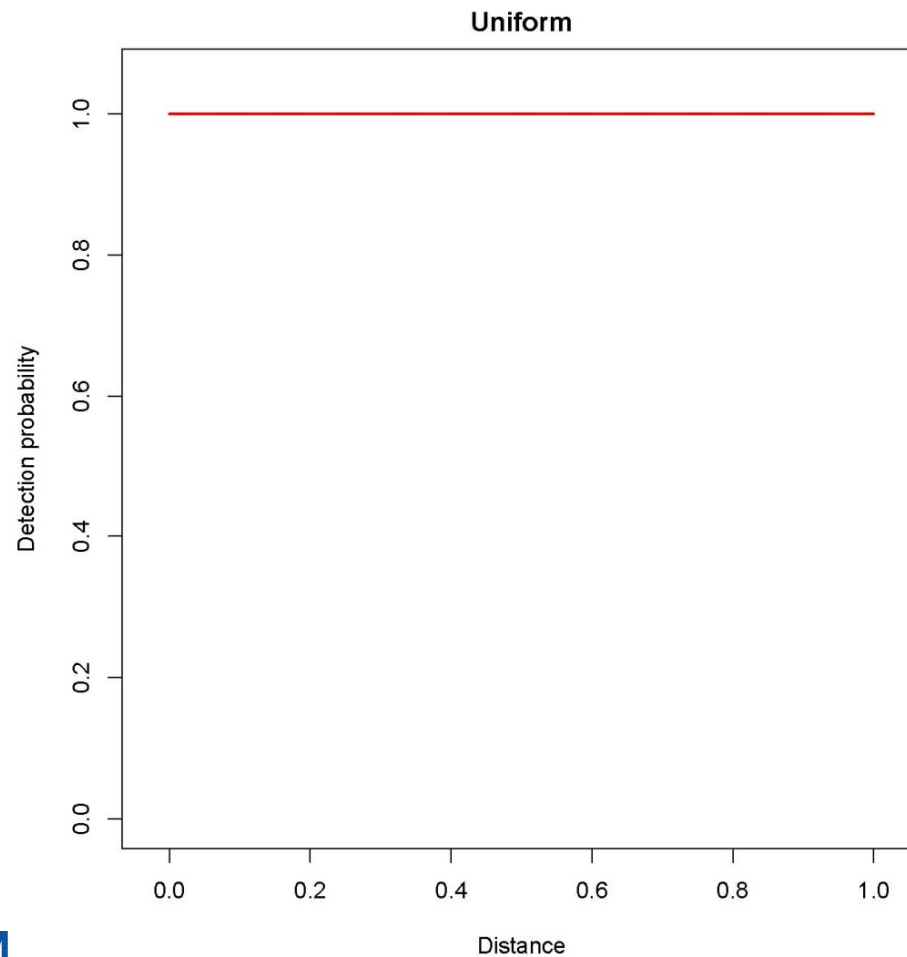
The first step in constructing a model for $g(x)$ is to choose a **key function**

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:

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

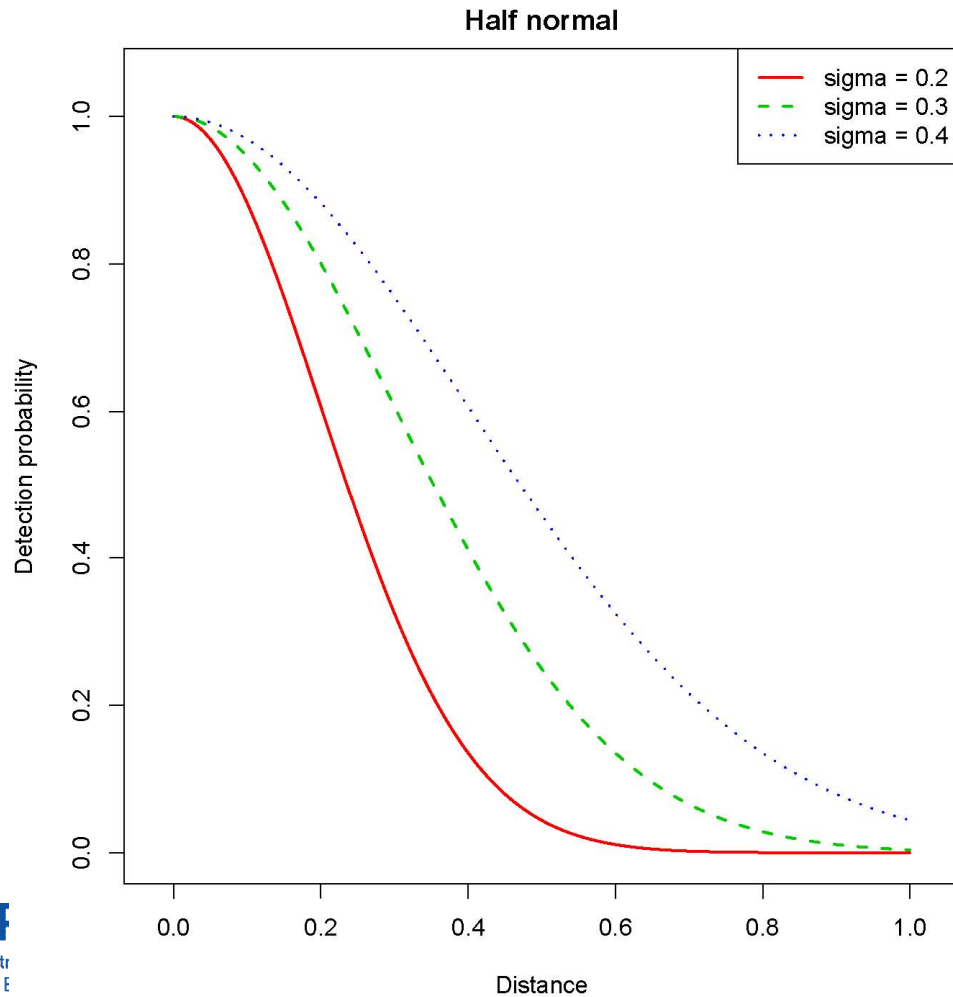
- Parameters = 0
- Shape criterion?

Yes

- Model robust?

No

Key functions (cont.)



- Model formula:

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

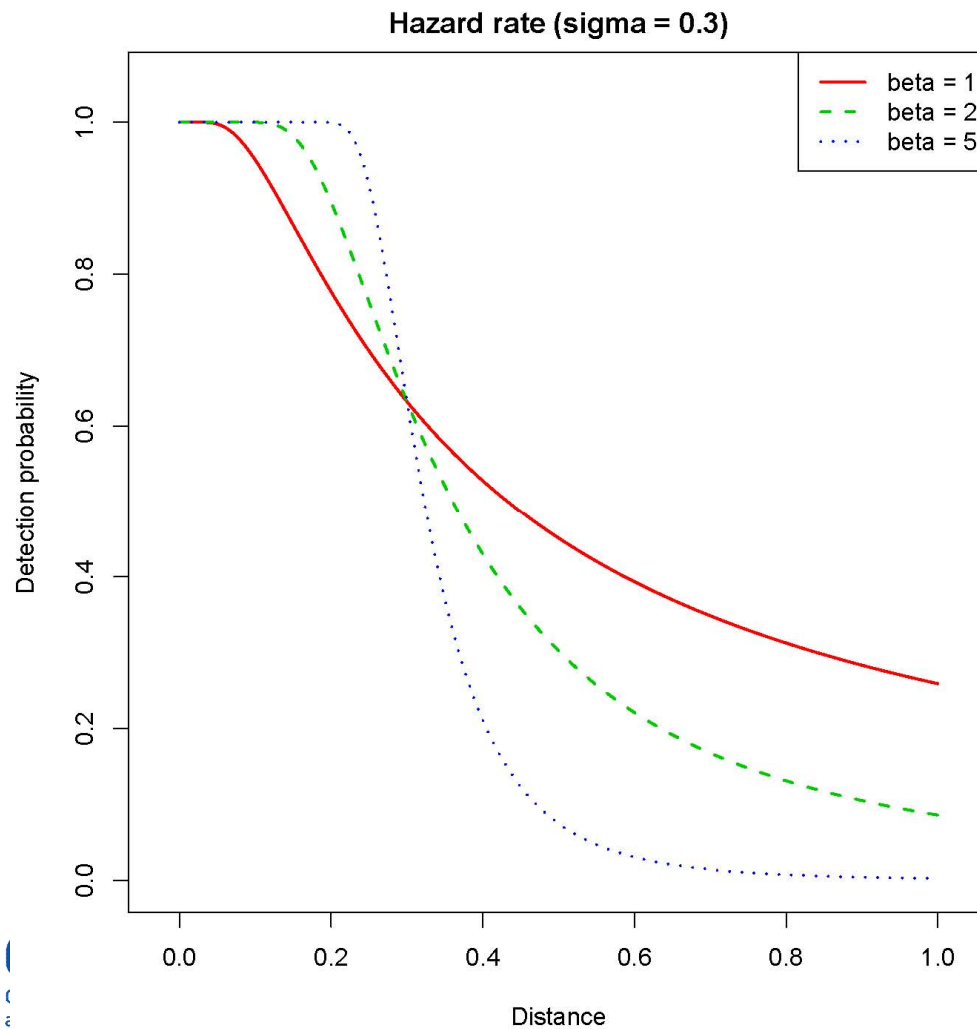
- Parameters = 1
- Shape criterion?

Yes

- Model robust?

No

Key functions (cont.)



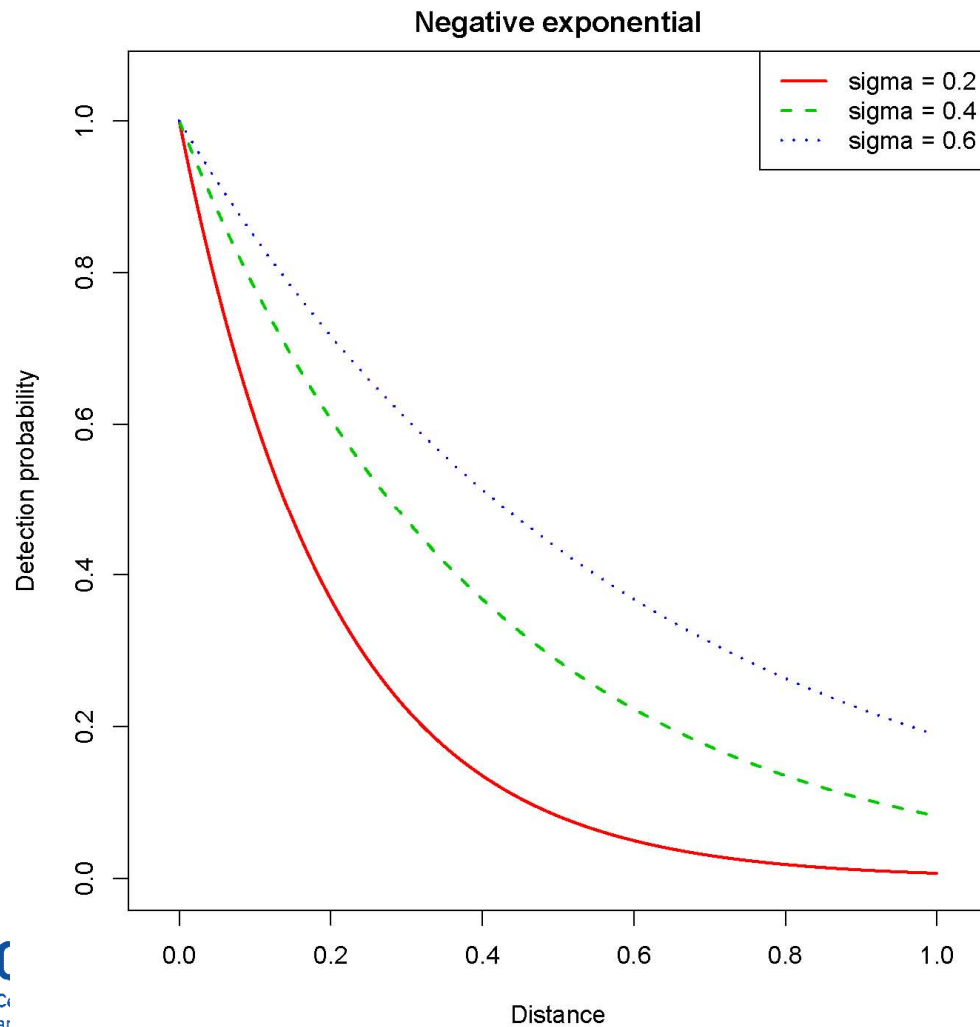
- Model formula:
$$g(x) = 1 - \exp\left[-\left(\frac{x}{\sigma}\right)^{-\beta}\right], x \leq w$$

- Parameters = 2
- Shape criterion?

Yes

Yes

Key functions (cont.)



- Model formula:

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

- Parameters = 1
- Shape criterion?
- Model robust?

No

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. A red arrow points from this button to the 'Key function' dropdown in the 'Model Definition Properties' dialog box on the right. The dialog box shows the 'Key function' set to 'Half-normal' and the 'Series expansion' set to 'Cosine'. The 'Analysis Engine' is set to 'CDS - Conventional distance sampling'.

Analysis 139: [a1 - HN by strat f0 pooled w82.5] Set: [AUK paper repo...]

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 Details ...

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

Comments

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

Defaults Name: HN by strat f0 pooled OK Cancel

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 $\times (1 + \text{Series})$

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

The α_i parameters must be estimated

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

The number of adjustment terms needs to be chosen

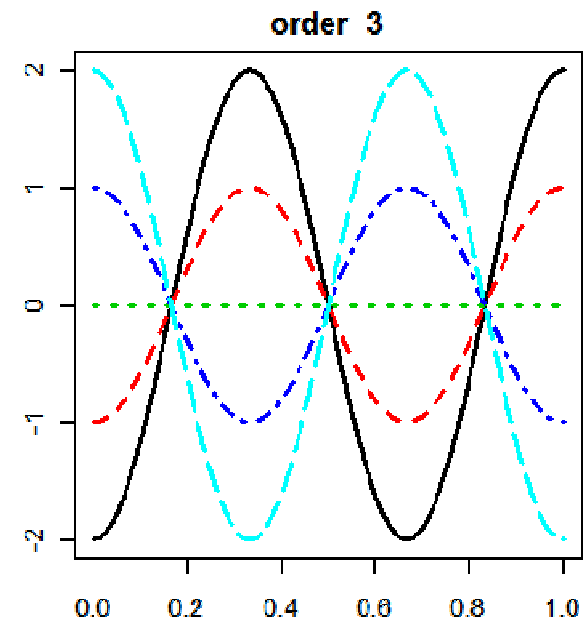
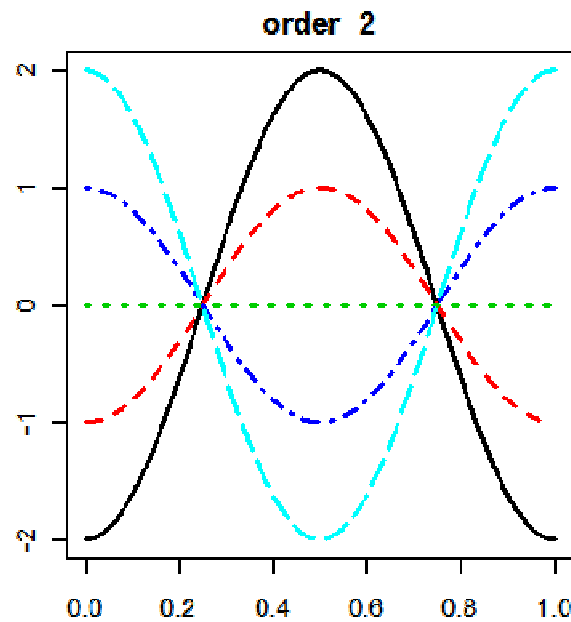
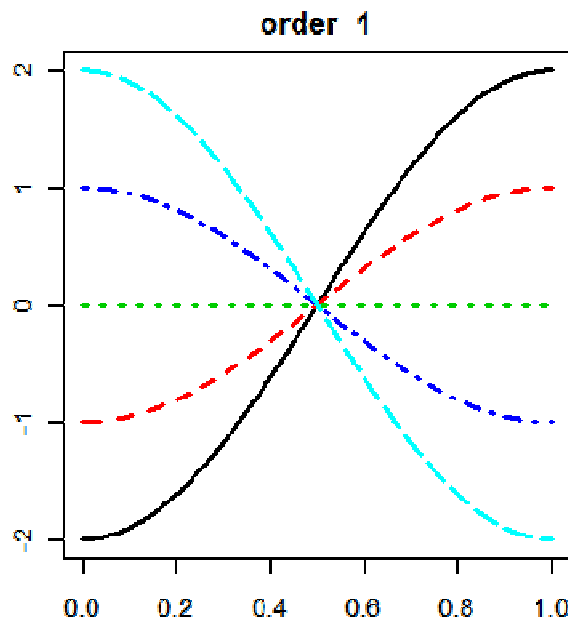
Adjustment terms

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

Key function	Series adjustment
Uniform*	Cosine*
Half normal [†]	Hermite polynomial [†]
Hazard rate	Simple polynomial
Negative exponential	

How adjustment terms work

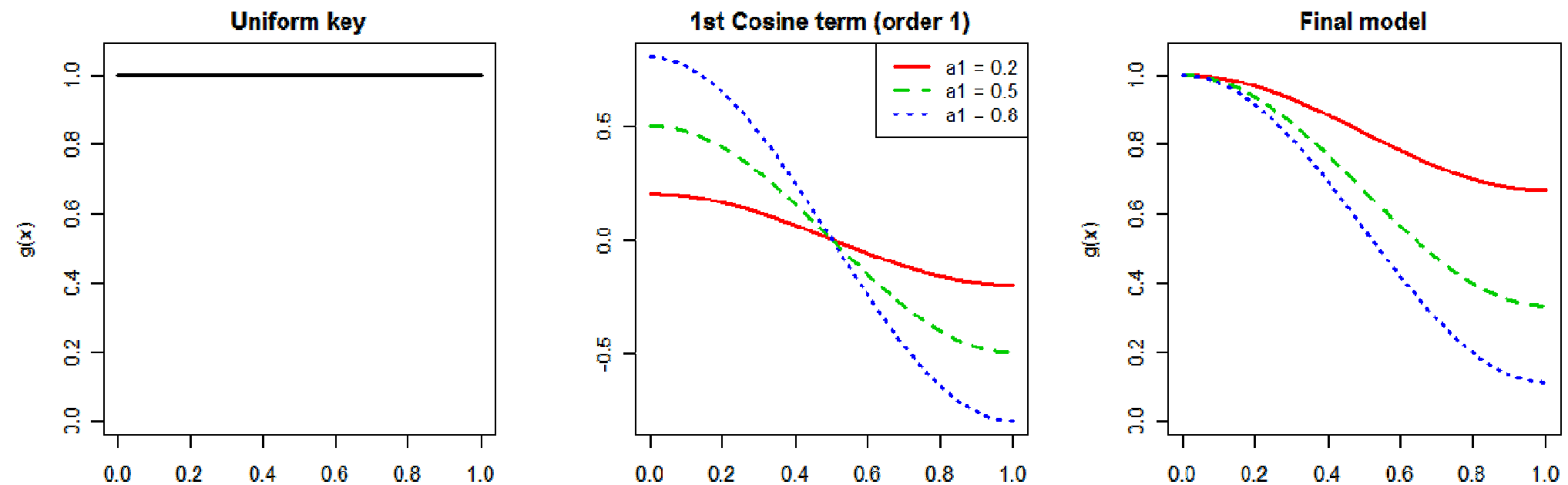
E.g. Cosine series (for different values of α)



(1st order only used for uniform)

How adjustment terms work

E.g. Uniform + 1 Cosine adjustment term:

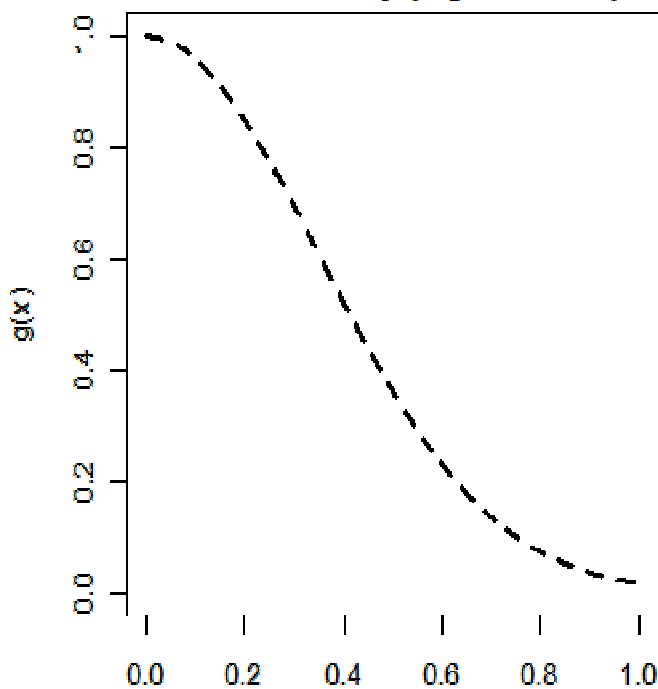


The effect of the adjustment terms depends on the value of their parameters

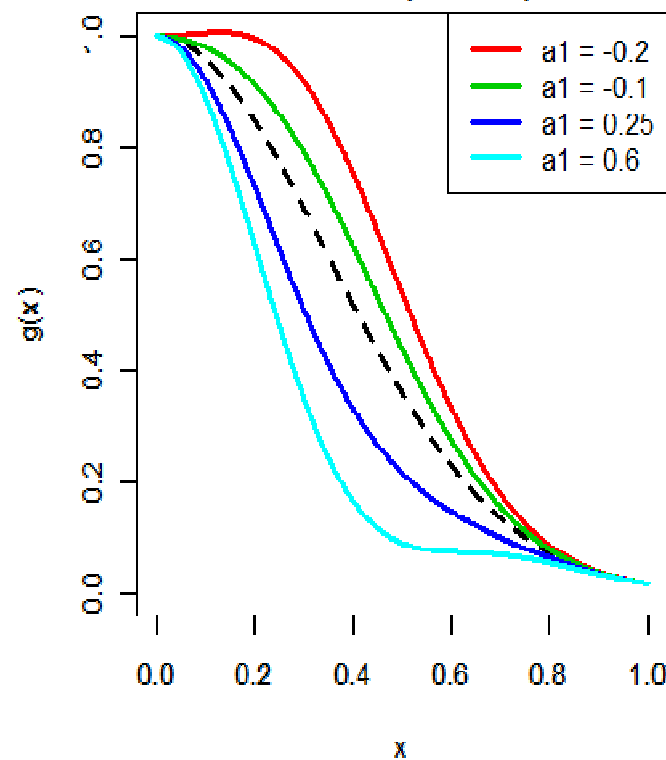
How adjustment terms work

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

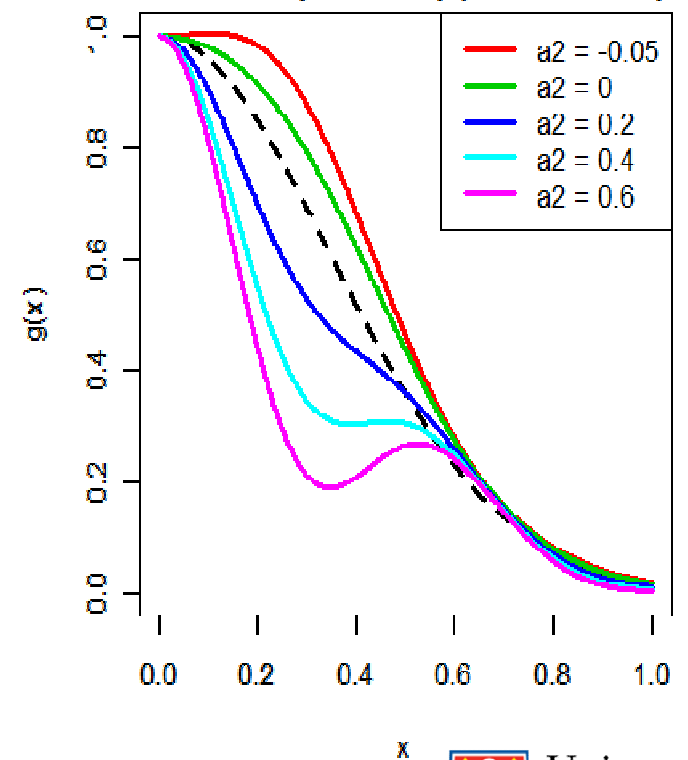
Half normal key ($\sigma = 0.35$)



+ 1 term (order 2)



+ 2 terms ($a_1 = -0.1$) (orders 2 & 3)



Adjustments in Distance

The screenshot displays the Distance software interface, divided into two main panels: 'Inputs' on the left and 'Results' on the right. The 'Inputs' panel shows the analysis name 'a1 - HN by strat f0 pooled w82.5', creation and run dates, survey details, data filters, and a list of model definitions. The 'Results' panel shows the 'Model Definition Properties' for the selected model, including the analysis engine, tabs for various settings, and a table of detection function models.

Inputs Panel:

- 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 list:
 - 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

Results Panel: Model Definition Properties: [HN by strat f0 pooled]

- Analysis Engine: CDS - Conventional distance sampling
- Tabs: Estimate, Detection function, Cluster size, Multipliers, Variance, Misc.
- Sub-tabs: Models, Adjustment terms, Constraints, Diagnostics
- Detection function models table:

Model	Key function	Series expansion
1	Half-normal	Cosine
- Select among multiple models using: AIC

Adjustments in Distance

The screenshot displays the Distance software interface. The main window shows the 'Analysis 139: [a1 - HN by strat f0 pooled w82.5] Set: [AUK paper repo...]' with fields for Name, Created, and Run. The 'Survey' section shows 'Set 1' and '[1] New Survey'. The 'Data filter' section lists '1 Default Data Filter' and '2 Trunc 82.5'. The 'Model definition' section lists several models, with '5 HN by strat f0 pooled' selected. A red circle highlights the 'Properties ...' button next to this model. An orange arrow points from this button to the 'Model Definition Properties: [HN by strat f0 pooled]' dialog box.

The 'Model Definition Properties' dialog box is open, showing the 'Adjustment terms' tab. The 'Analysis Engine' is set to 'CDS - Conventional distance sampling'. The 'Estimate' tab is selected, and the 'Adjustment terms' sub-tab is active. The 'Selection of adjustment terms' section shows 'Automated selection' chosen. The 'Selection method' is 'Sequential', 'Look-ahead' is '1', 'Selection criterion' is 'AIC', 'Significance level' is '0.15', and 'Maximum terms' is '5'. The 'Manual selection' section is empty. The 'Scaling of distances' section shows 'Scale distances by: W (right truncation distance)'. The 'Name' field at the bottom is 'HN 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

Selection of adjustment terms

☒ Automated selection

Selection method: Sequential Look-ahead: 1

Selection criterion: AIC Significance level: 0.15

Maximum terms: 5

☐ Manual selection

Model	Num adj. parameters	Order of adjustment parameters (optional)
1	0	

☐ Manually select starting values

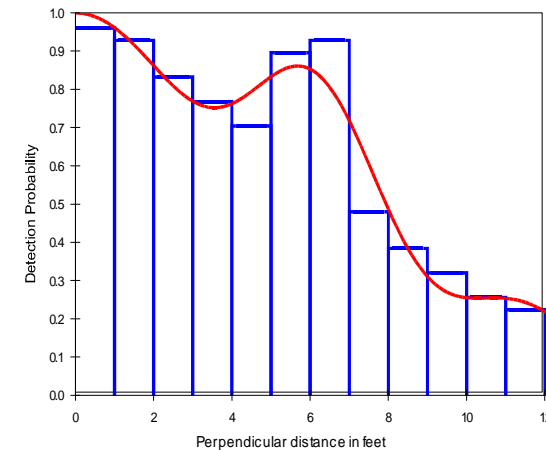
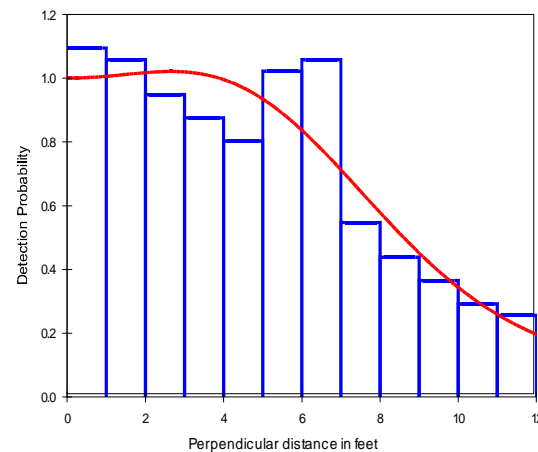
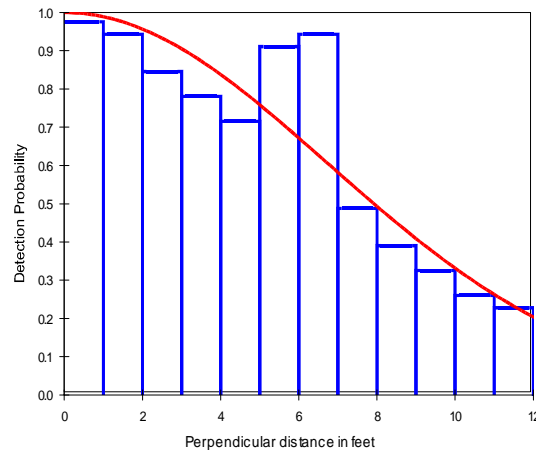
Model	Num parameters
1	0

Scaling of distances

Scale distances by: W (right truncation distance)

Defaults Name: HN by strat f0 pooled OK Cancel

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

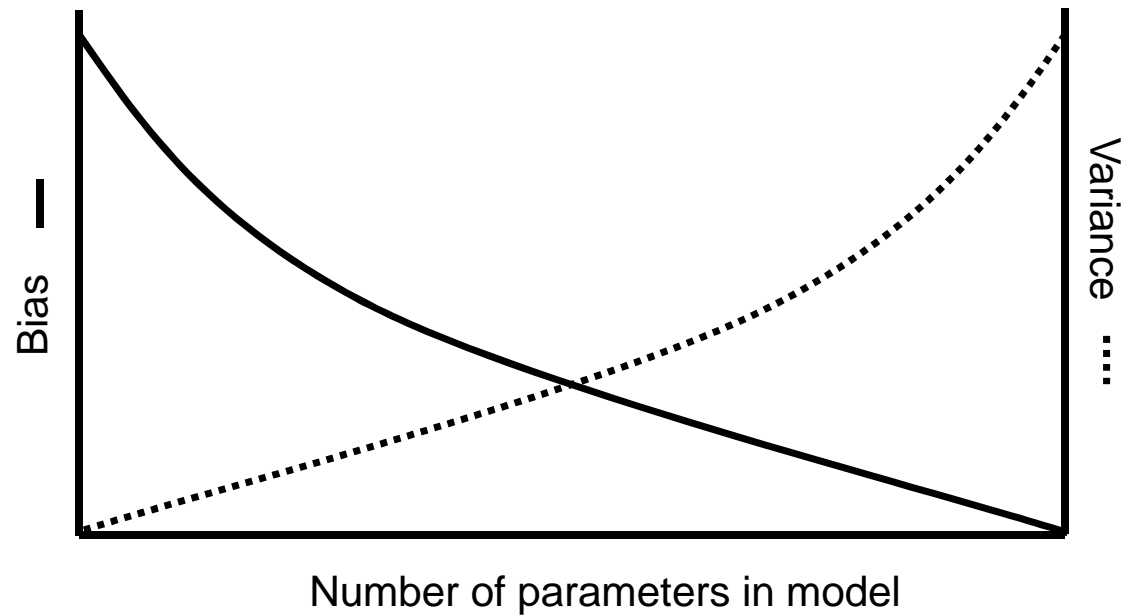
How many parameters?

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