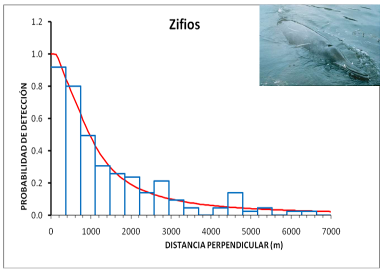
13. Line transect data analysis



$$\hat{D} = \frac{n}{2wL\hat{p}_a}$$

$$\hat{N} = \hat{D}\,A$$

Abundance estimation - résumé

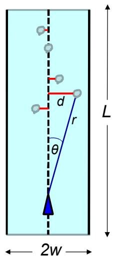
- In a strip transect we assume we see everything

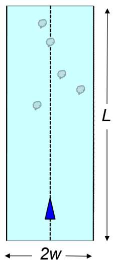
$$\hat{D} = \frac{n}{a} = \frac{n}{2wL}$$

- In line transect sampling, not everything is seen
  - Need to estimate the probability of detection, p_a
  - p_a is estimated from perpendicular distance data

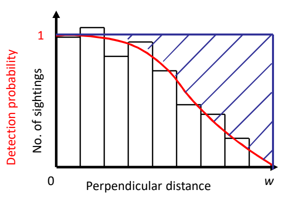
$$\hat{D} = \frac{n}{2wL\hat{p}_a}$$

$$\hat{N} = \hat{D}\,A$$





Estimating probability of detection, p_a - résumé



Average probability of detection: $\hat{p}_a = \dfrac{\text{area under curve}}{\text{area under rectangle}}$

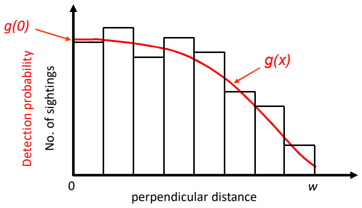
Important assumption: animals are distributed uniformly between 0 and w

Line transect sampling assumptions - résumé

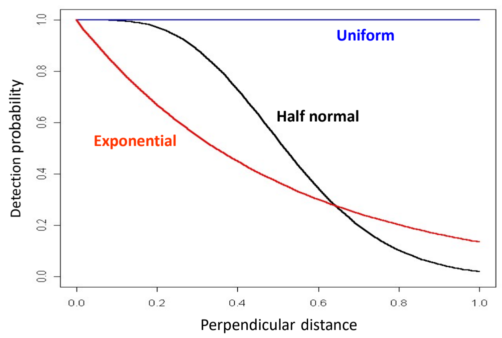
- All measurements are accurate
- All animals on the transect line are detected
  - Detection probability is 1 at zero perpendicular distance
- All animals are detected at their initial location
  - No movement before detection
    - especially in response to survey platform
- Sample data are representative
- Observations are independent
  - For unbiased variance estimation

Criteria for a good detection function

- A model that:
  - Can fit a wide variety of plausible shapes (model robust)
  - Can fit data when many factors affect detectability (pooling robust)
  - Has a shoulder: $g'(0)=0$ (shape)
  - Generates estimates with good precision (efficiency)



Some detection functions



Detection function models

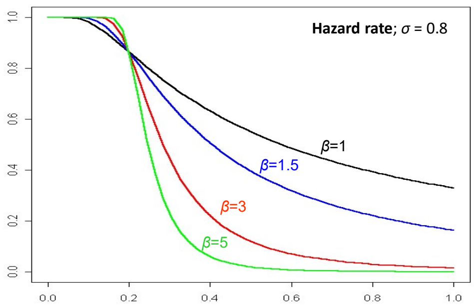
Uniform: $g(x) = 1, \; x \le w$

Half-normal: $g(x) = exp\left(\frac{-x^2}{2\sigma^2}\right), \; x \le w$

Hazard-rate: $g(x) = 1 - exp\left[-\left(\frac{x}{\sigma}\right)^{-\beta}\right], \; x \le w$

x = perpendicular distance
w = truncation distance (width of nominal strip)
σ = standard deviation of perpendicular distance
β = parameter to be estimated

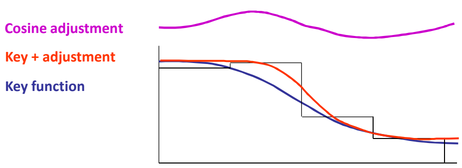
Hazard rate detection function



Adjustment terms to improve robustness

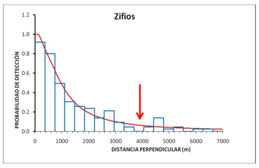
- Series adjustments to the "key" detection function:
  $g(x)$ = *key function* × (1 + *series adjustment*)
- Series adjustments
  - Cosine
  - Simple polynomial
  - Hermite polynomial

  $\Big\}$ All "wavy" functions



Data truncation

- Sightings at large perpendicular distance contribute little to estimation
- But may lead to poor model fit and high variance



- Consider truncating data before model fitting
  - Not uncommon to truncate around 5% of observations
  - But need to decide on a case-by-case basis
    - Might be better to choose a particular distance

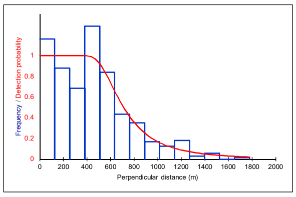
Akaike's Information Criterion

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

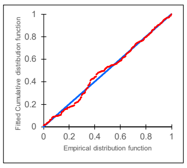
- ***L*** is the maximized likelihood
  - The likelihood is a statistical function relating the parameters of a model to the data
- ***q*** is the number of model parameters
- Used for comparing models
- Model with smallest AIC has most support from the data
  - If two models have a difference in AIC of less than 2 units, they are considered to have more or less equivalent support from the data
- Can only use when the data are the same
  - So, cannot compare models with different truncation distances

Goodness of fit tests: how well does the model really fit the data?

- Chi-squared test
  - For grouped data
  - ∑ [(Obs - Exp)2 / Exp]
  - Not very useful because the result depends on how the data are grouped

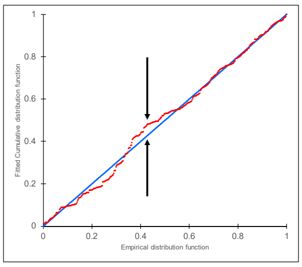


- More useful is the Q-Q plot
  - Quantile-Quantile plot
    - Shows fit to data point-by-point
  - Also
    - Kolmogorov-Smirnov (KS) test
    - Cramér-von-Mises (CvM) tests



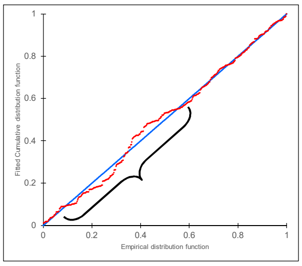
Kolmogorov-Smirnov test

- Uses largest difference between lines to measure goodness of fit



Cramér-von-Mises test

- Uses the sum of all squared differences to measure goodness of fit
  - Can also use a weighted sum, which is a better measure of fit close to the transect line (good)



Goodness of fit tests - summary

- Chi-squared test requires data to be grouped
  - Results depend on grouping
- Q-Q plots show goodness of fit at “high resolution”
  - without requiring grouping into intervals
- Kolmogorov-Smirnov and Cramér-von-Mises tests
  - also without grouping into intervals
- Cramér-von-Mises test can be weighted
  - to give higher weight to observations near zero

Group size

- If targets occur in groups (schools)
  - Estimate density of groups, D_s
  - Calculate mean group size, $\bar{s}$
  - Multiply to get density of animals, D

$$\hat{D} = \hat{D}_s \bar{s} \qquad \hat{D}_s = \frac{n}{2wL\hat{P}_a}$$

n = number of groups

- BUT smaller groups are more likely to be missed at greater perpendicular distance
  - Regress group size (or log group size) on perpendicular distance (or detection probability)
  - Value at intercept (distance = 0) should be unbiased
  - In this case, we use: $\hat{D} = \hat{D}_s E[s]$
    - Expected (estimated) group size, E[*s*], rather than mean group size

Results

- Parameter estimates (for species in groups)

  p_a = average detection probability within truncation distance, *w*

  esw = effective strip width = $p_a \times w$

  f(0) = 1 / *esw*

  n/L = encounter rate of groups

  D_s = density of groups

  E[s] = expected group size

  D = density of animals

  N = abundance of animals

Summary

- Estimate detection function
  - Model
    - Key function
    - Adjustment terms
  - Consider truncating data
  - Model selection
    - AIC
    - Goodness of fit
  - Group size correction