Estimation of Finite Mixtures with Nonparametric Components

C. S. Chee

Department of Statistics
The University of Auckland
chee@stat.auckland.ac.nz

Joint work with Dr Yong Wang
1 Introduction
- Preamble
- The Kernel-based Semiparametric Model
- The Mixture-based Semiparametric Model

2 Estimation and Selection
- Model Parameter Estimation
- Tuning Parameter Selection

3 Real Data Examples

4 Simulation
- Data
- Performance Measures
- Results

5 Summary
Part I
General Idea

Consider a mixture distribution having density $g(x)$:

$$g(x) = \sum_{k=1}^{K} \lambda_k f(x - \mu_k)$$

Focus: A two-component mixture of location-shifted distributions ($K = 2$).

Goal: Estimate $\lambda_1, \mu_1, \mu_2$ and $f$ given an iid sample $\{x_i\}_{1 \leq i \leq n}$ from $g(x)$.

Approach: Use a nonparametric mixture for $f$. 
A Running Example

- This dataset contains the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser.
- A two-component mixture model is reasonable.
Identifiability Result

Model

\[ g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2) \]

- Do not want to assume \( f \) belongs to a parametric family.
- **Question**: Is \( g(x) \) identifiable?
- \( g(x) \) is identifiable if it is unique as a function of \( \lambda, \mu_1, \mu_2 \) and \( f \).
**Identifiability Result**

**Model**

\[ g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2) \]

- Do not want to assume \( f \) belongs to a parametric family.
- **Question**: Is \( g(x) \) identifiable?
- \( g(x) \) is identifiable if it is unique as a function of \( \lambda, \mu_1, \mu_2 \) and \( f \).
- Bordes *et al.* (2006) and Hunter *et al.* (2007) showed that \( g(x) \) is identifiable if \( f \) is a symmetric density about zero when \( \lambda \neq \frac{1}{2} \) and \( \mu_1 \neq \mu_2 \).
The Kernel-based Semiparametric Model

Bordes et al. (2007)

\[ g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2) \]

with \( f \) being the symmetrized nonparametric kernel given by

\[ f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left( \frac{y - x_i + \mu_{z_i}}{h} \right) + K\left( \frac{y + x_i - \mu_{z_i}}{h} \right) \right\}, \]

where

- \( K(\cdot) \) — a kernel density function,
- \( h > 0 \) — the bandwidth,
- \( z_i \in \{1, 2\} \) — the component label of \( x_i \).
Construction of $g$

\[ g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2) \]

\[ f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left( \frac{y - (x_i - \mu_{zi})}{h} \right) + K\left( \frac{y - [- (x_i - \mu_{zi})]}{h} \right) \right\} \]

For illustration, consider $h = 1$, $n = 2$ and $K(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$.

Assume that $z_i$ and $\mu_{zi}$ are known.

Let $x_i^* = x_i - \mu_{zi}$.
Construction of $g$

$$f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left(\frac{y - x_i^*}{h}\right) + K\left(\frac{y - (-x_i^*)}{h}\right) \right\}$$
Construction of $g$

\[ f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left( \frac{y - x_i^*}{h} \right) + K\left( \frac{y - (-x_i^*)}{h} \right) \right\} \]
Construction of $g$

$$f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left( \frac{y - x_i^*}{h} \right) + K\left( \frac{y - (-x_i^*)}{h} \right) \right\}$$
Construction of $g$

$$f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left( \frac{y - x_i^*}{h} \right) + K\left( \frac{y - (-x_i^*)}{h} \right) \right\}$$
Construction of $g$

$$f(y) = \sum_{i=1}^{n} \frac{1}{2nh} \left\{ K\left(\frac{y - x_i^*}{h}\right) + K\left(\frac{y - (-x_i^*)}{h}\right) \right\}$$
Construction of $g$

$$g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2)$$
Construction of $g$

$$g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2)$$
Construction of $g$

$$g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2)$$
Old Faithful Geyser Data

$h = 3.99$
Old Faithful Geyser Data

\[
\frac{\lambda}{2nh} \times \left\{ \begin{array}{l}
\hat{p}_{11}K\left( \frac{x - \hat{\mu}_1 - x_1 + \hat{\mu}_1}{h} \right) + \hat{p}_{11}K\left( \frac{x - \hat{\mu}_1 + x_1 - \hat{\mu}_1}{h} \right) + \hat{p}_{12}K\left( \frac{x - \hat{\mu}_1 - x_1 + \hat{\mu}_2}{h} \right) + \hat{p}_{12}K\left( \frac{x - \hat{\mu}_1 + x_1 - \hat{\mu}_2}{h} \right) \\
+ \hat{p}_{21}K\left( \frac{x - \hat{\mu}_1 - x_2 + \hat{\mu}_1}{h} \right) + \hat{p}_{21}K\left( \frac{x - \hat{\mu}_1 + x_2 - \hat{\mu}_1}{h} \right) + \hat{p}_{22}K\left( \frac{x - \hat{\mu}_1 - x_2 + \hat{\mu}_2}{h} \right) + \hat{p}_{22}K\left( \frac{x - \hat{\mu}_1 + x_2 - \hat{\mu}_2}{h} \right) \\
+ \cdots \\
+ \hat{p}_{n1}K\left( \frac{x - \hat{\mu}_1 - x_n + \hat{\mu}_1}{h} \right) + \hat{p}_{n1}K\left( \frac{x - \hat{\mu}_1 + x_n - \hat{\mu}_1}{h} \right) + \hat{p}_{n2}K\left( \frac{x - \hat{\mu}_1 - x_n + \hat{\mu}_2}{h} \right) + \hat{p}_{n2}K\left( \frac{x - \hat{\mu}_1 + x_n - \hat{\mu}_2}{h} \right) \end{array} \right\}
\]

\[
\frac{(1 - \lambda)}{2nh} \times \left\{ \begin{array}{l}
\hat{p}_{11}K\left( \frac{x - \hat{\mu}_2 - x_1 + \hat{\mu}_1}{h} \right) + \hat{p}_{11}K\left( \frac{x - \hat{\mu}_2 + x_1 - \hat{\mu}_1}{h} \right) + \hat{p}_{12}K\left( \frac{x - \hat{\mu}_2 - x_1 + \hat{\mu}_2}{h} \right) + \hat{p}_{12}K\left( \frac{x - \hat{\mu}_2 + x_1 - \hat{\mu}_2}{h} \right) \\
+ \hat{p}_{21}K\left( \frac{x - \hat{\mu}_2 - x_2 + \hat{\mu}_1}{h} \right) + \hat{p}_{21}K\left( \frac{x - \hat{\mu}_2 + x_2 - \hat{\mu}_1}{h} \right) + \hat{p}_{22}K\left( \frac{x - \hat{\mu}_2 - x_2 + \hat{\mu}_2}{h} \right) + \hat{p}_{22}K\left( \frac{x - \hat{\mu}_2 + x_2 - \hat{\mu}_2}{h} \right) \\
+ \cdots \\
+ \hat{p}_{n1}K\left( \frac{x - \hat{\mu}_2 - x_n + \hat{\mu}_1}{h} \right) + \hat{p}_{n1}K\left( \frac{x - \hat{\mu}_2 + x_n - \hat{\mu}_1}{h} \right) + \hat{p}_{n2}K\left( \frac{x - \hat{\mu}_2 - x_n + \hat{\mu}_2}{h} \right) + \hat{p}_{n2}K\left( \frac{x - \hat{\mu}_2 + x_n - \hat{\mu}_2}{h} \right) \end{array} \right\}
\]

C. S. Chee
Semiparametric Mixtures of Mixture
The Mixture-based Semiparametric Model

Our proposal

\[ g(x) = \lambda f(x - \mu_1) + (1 - \lambda) f(x - \mu_2) \]

with \( f \) being the symmetrized nonparametric mixture given by

\[
f(y; \pi, \theta) = \sum_{j=1}^{m} \frac{\pi_j}{2h} \left\{ \phi\left(\frac{y - \theta_j}{h}\right) + \phi\left(\frac{y + \theta_j}{h}\right) \right\},
\]

where

- \( \phi(\cdot) \) — a known unimodal density that is symmetric about zero,
- \( h > 0 \) — the (known or fixed) tuning parameter,
- \( \theta = (\theta_1, \ldots, \theta_m)^\top \) — a support point vector,
- \( \pi = (\pi_1, \ldots, \pi_m)^\top \) — the corresponding probability mass vector.
Old Faithful Geyser Data

$h = 5.9$
A Comparison

![Comparison of Kernel and Mixture Models](chart.png)
A Comparison

**Kernel** (left component)

\[
\frac{\hat{\lambda}}{2nh} \times \left\{ \begin{array}{c}
\hat{p}_{11} K \left( \frac{x - \hat{\mu}_1 - x_1 + \hat{\mu}_1}{h} \right) + \hat{p}_{11} K \left( \frac{x - \hat{\mu}_1 + x_1 - \hat{\mu}_1}{h} \right) + \hat{p}_{12} K \left( \frac{x - \hat{\mu}_1 - x_1 + \hat{\mu}_2}{h} \right) + \hat{p}_{12} K \left( \frac{x - \hat{\mu}_1 + x_1 - \hat{\mu}_2}{h} \right) \\
+ \hat{p}_{21} K \left( \frac{x - \hat{\mu}_1 - x_2 + \hat{\mu}_1}{h} \right) + \hat{p}_{21} K \left( \frac{x - \hat{\mu}_1 + x_2 - \hat{\mu}_1}{h} \right) + \hat{p}_{22} K \left( \frac{x - \hat{\mu}_1 - x_2 + \hat{\mu}_2}{h} \right) + \hat{p}_{22} K \left( \frac{x - \hat{\mu}_1 + x_2 - \hat{\mu}_2}{h} \right) \\
+ \cdots \\
+ \hat{p}_{n1} K \left( \frac{x - \hat{\mu}_1 - x_n + \hat{\mu}_1}{h} \right) + \hat{p}_{n1} K \left( \frac{x - \hat{\mu}_1 + x_n - \hat{\mu}_1}{h} \right) + \hat{p}_{n2} K \left( \frac{x - \hat{\mu}_1 - x_n + \hat{\mu}_2}{h} \right) + \hat{p}_{n2} K \left( \frac{x - \hat{\mu}_1 + x_n - \hat{\mu}_2}{h} \right) \end{array} \right\}
\]

**Mixture** (left component)

\[
\frac{\hat{\lambda}}{2h} \times \left\{ \phi \left( \frac{x - \hat{\mu}_1 - \hat{\theta}}{h} \right) + \phi \left( \frac{x - \hat{\mu}_1 + \hat{\theta}}{h} \right) \right\}
\]
Part II
Model Parameter Estimation

- Assume that \( h \) is **KNOWN**.
- Employ maximum likelihood estimation of the model parameters.
- Denote by \( G \) a discrete distribution formed by the \( m \) points of support \( \theta \) with corresponding masses \( \pi \).
- Let \( \beta = (\lambda, \mu_1, \mu_2)^\top \).
- The log-likelihood function: \( \ell_h(G, \beta) = \sum_{i=1}^{n} \log g_h(x_i; G, \beta) \)
- \( \hat{G} \) and \( \hat{\beta} \) (estimates of \( G \) and \( \beta \)) can be found by the algorithm of Wang (2009).
Model Parameter Estimation

- Assume that \( h \) is **KNOWN**.
- Employ maximum likelihood estimation of the model parameters.
- Denote by \( G \) a discrete distribution formed by the \( m \) points of support \( \theta \) with corresponding masses \( \pi \).
- Let \( \beta = (\lambda, \mu_1, \mu_2)^\top \).
- The log-likelihood function: \( \ell_h(G, \beta) = \sum_{i=1}^{n} \log g_h(x_i; G, \beta) \)
- \( \hat{G} \) and \( \hat{\beta} \) (estimates of \( G \) and \( \beta \)) can be found by the algorithm of Wang (2009).
- What if \( h \) is **UNKNOWN**?
Tuning Parameter Selection

- Select a “good” $h$ from a set of predetermined candidates.
- General approaches to model selection can be used:
  - cross-validation (CV)
  - information criteria
- General strategy for automatic selection of tuning parameter:
  - **Step 1**: Choose a selection criterion.
  - **Step 2**: Compute the value of the specified selection criterion over a grid of tuning parameters.
  - **Step 3**: Select the $h$ that has the minimum value of the selection criterion.
Cross-Validation

- Two CV-based criteria:
  - LSCV — the least-squares cross-validation criterion
  - LCV — the likelihood cross-validation criterion

- One of the commonly used CV methodologies is the $V$-fold CV.

- We set $V = 10$. 
Illustration: $V$-fold CV

$\{x_i\}_{1 \leq i \leq n}$ is split into $V$ roughly equal-sized and non-overlapping subsets $S_1, \ldots, S_V$.

- Select $S_1$ as the test set and the remaining folds as the training set.
- Model parameters are estimated based on the training set.
- Fitted model is evaluated on the test set.
- Repeat for $S_2, \ldots, S_V$. 

![Diagram of V-fold CV](image-url)
CV-based Criteria

The LSCV and LCV criteria implemented via the $V$-fold CV procedure are respectively defined by

\[
LSCV(h) = \frac{1}{V} \sum_{v=1}^{V} \int \left\{ \hat{g}_{-v}(x; \hat{G}, \hat{\beta}, h) \right\}^2 dx - \frac{2}{V} \sum_{v=1}^{V} \sum_{x_j \in S_v} \frac{1}{|S_v|} \hat{g}_{-v}(x_j; \hat{G}, \hat{\beta}, h)
\]

and

\[
LCV(h) = -\frac{1}{V} \sum_{v=1}^{V} \sum_{x_j \in S_v} \frac{1}{|S_v|} \log \hat{g}_{-v}(x_j; \hat{G}, \hat{\beta}, h),
\]

where $|S_v|$ denotes the cardinality of $S_v$ and $\hat{g}_{-v}(x; \hat{G}, \hat{\beta}, h)$ is the fitted model based on all the data points except the observations belonging to the subset $S_v$. 
Information Criteria

- Two popular information criteria:

\[
\begin{align*}
\text{AIC}(h) &= -2\ell_h(\hat{G}, \hat{\beta}) + 2p \\
\text{BIC}(h) &= -2\ell_h(\hat{G}, \hat{\beta}) + p \log(n)
\end{align*}
\]

where \( p \) is the number of free parameters.

- We can also use a small sample version of AIC, called \( \text{AIC}_c \) (see Burnham and Anderson, 2002):

\[
\text{AIC}_c(h) = \text{AIC}(h) + \frac{2p(p+1)}{n-p-1}
\]
Part III
Real Data Examples

- Compare our mixture-based semiparametric model against the kernel-based semiparametric model.
- Used the algorithm of Benaglia et al. (2009) for fitting the kernel-based model.
- \( K(\cdot) \) and \( \phi(\cdot) \) were taken to be the standard Gaussian density.
  - Example 1: 2008 World Fly Fishing Championships Data
  - Example 2: Exploring Relationships in Body Dimensions
  - Example 3: Australian Athletes Data
Example 1: 2008 World Fly Fishing Championships Data

- The 2008 WFFC was held in the Taupo-Rotorua regions; details may be obtained at Yee (2009).
- Considered the length of fish caught in Lake Rotoaira \((n = 201)\).
Result 1: Fish Length Data

(a) Kernel (LSCV)  
(2.25, 0.13, 25.91, 46.41)

(b) Kernel (LCV)  
(1.75, 0.13, 25.93, 46.47)

(c) Mixture (AIC<sub>c</sub>)  
(5.1, 0.14, 26.04, 46.54)

(d) Mixture (LCV)  
(5.2, 0.13, 26.02, 46.53)
Example 2: Exploring Relationships in Body Dimensions

- 21 body dimension measurements as well as age, weight, height and gender on 247 men and 260 women (Heinz et al., 2003).
- Used the variable elbow diameter.
Result 2: Elbow Diameter Data

LCV plot for Kernel

- $h = 0.05$
- $h = 0.21$
- $h = 0.32$

C. S. Chee  
Semiparametric Mixtures of Mixtures
Result 2: Elbow Diameter Data

(a) Kernel (LSCV)  
(0.29, 0.56, 12.57, 14.43)

(b) Kernel (LCV)  
(0.32, 0.48, 13.39, 13.39)

(c) Mixture (AIC<sub>c</sub>)  
(0.87, 0.56, 12.47, 14.56)

(d) Mixture (LCV)  
(0.87, 0.56, 12.47, 14.56)
Example 3: Australian Athletes Data

- Data on 102 male and 100 female Australian athletes collected at the Australian Institute of Sport (Cook and Weisberg, 1994).
- Used the variable lean body mass (LBM).
Result 3: LBM Data

(a) Kernel (LSCV)  
(3.25, 0.87, 64.87, 64.87)

(b) Kernel (LCV)  
(3.25, 0.87, 64.87, 64.87)

(c) Mixture (AIC_c)  
(7.2, 0.63, 57.37, 77.3)

(d) Mixture (LCV)  
(8, 0.63, 57.44, 77.21)
Part IV
Consider three two-component mixtures (NOmix, T5mix, PEmix) of these components:

- standard normal (NO)
- t-distribution with 5 dof (T5)
- standard power exponential with shape parameter \( \nu = 4 \) (PE)

with \( \lambda = 0.3, \mu_1 = -2, \mu_2 = 2 \) and \( \sigma = 1 \).

In each simulation example, 100 replications with \( n = 200 \) were generated.
Performance Measures

- To select $h$:
  
  LSCV and LCV

- To assess parameter estimates:
  
  $$SE(\alpha, \hat{\alpha}) = (\hat{\alpha} - \alpha)^2$$

- To assess density estimates:
  
  $$ISE(g, \hat{g}) = \int [\hat{g}(x) - g(x)]^2 \, dx$$
  $$KLL(g, \hat{g}) = \int g(x) \log \frac{g(x)}{\hat{g}(x)} \, dx$$
### Result (a): Parameter Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>Selection Method</th>
<th>PE\text{mix}</th>
<th>NO\text{mix}</th>
<th>T5\text{mix}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>[\widehat{\text{MSE}} \times 10^{-3}] of (\lambda) estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel</td>
<td>LSCV</td>
<td>0.96 (0.15)</td>
<td>1.26 (0.17)</td>
<td>1.58 (0.23)</td>
</tr>
<tr>
<td>Kernel</td>
<td>LCV</td>
<td>0.95 (0.15)</td>
<td>1.27 (0.17)</td>
<td>1.65 (0.23)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LSCV</td>
<td>0.92 (0.14)</td>
<td>1.23 (0.17)</td>
<td>1.43 (0.21)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LCV</td>
<td>0.92 (0.14)</td>
<td>1.22 (0.17)</td>
<td>1.45 (0.21)</td>
</tr>
<tr>
<td>(ii)</td>
<td>[\widehat{\text{MSE}} \times 10^{-2}] of (\mu_1) estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel</td>
<td>LSCV</td>
<td>2.55 (0.28)</td>
<td>2.32 (0.36)</td>
<td>5.02 (0.61)</td>
</tr>
<tr>
<td>Kernel</td>
<td>LCV</td>
<td>2.50 (0.28)</td>
<td>2.30 (0.36)</td>
<td>4.82 (0.59)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LSCV</td>
<td>2.30 (0.29)</td>
<td>2.31 (0.35)</td>
<td>3.97 (0.53)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LCV</td>
<td>2.10 (0.27)</td>
<td>2.33 (0.34)</td>
<td>4.00 (0.56)</td>
</tr>
<tr>
<td>(iii)</td>
<td>[\widehat{\text{MSE}} \times 10^{-2}] of (\mu_2) estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel</td>
<td>LSCV</td>
<td>1.11 (0.16)</td>
<td>0.87 (0.12)</td>
<td>1.30 (0.20)</td>
</tr>
<tr>
<td>Kernel</td>
<td>LCV</td>
<td>1.03 (0.15)</td>
<td>0.88 (0.12)</td>
<td>1.62 (0.28)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LSCV</td>
<td>1.03 (0.18)</td>
<td>0.87 (0.11)</td>
<td>1.13 (0.18)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LCV</td>
<td>0.86 (0.13)</td>
<td>0.87 (0.12)</td>
<td>1.06 (0.17)</td>
</tr>
</tbody>
</table>

**NOTE:** The values in parentheses are the corresponding standard errors.
## Result (b): Density Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>Selection Method</th>
<th>PEmix</th>
<th>NOmix</th>
<th>T5mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>( \widehat{\text{MISE}} ) ( \times 10^{-3} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel</td>
<td>LSCV</td>
<td>3.08 (0.18)</td>
<td>3.10 (0.20)</td>
<td>3.87 (0.24)</td>
</tr>
<tr>
<td>Kernel</td>
<td>LCV</td>
<td>3.22 (0.19)</td>
<td>2.87 (0.18)</td>
<td>4.04 (0.27)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LSCV</td>
<td>2.96 (0.20)</td>
<td>2.56 (0.18)</td>
<td>3.27 (0.22)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LCV</td>
<td>2.73 (0.18)</td>
<td>2.24 (0.16)</td>
<td>3.11 (0.21)</td>
</tr>
<tr>
<td>(ii)</td>
<td>( \widehat{\text{EKLL}} ) ( \times 10^{-2} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel</td>
<td>LSCV</td>
<td>2.13 (0.09)</td>
<td>1.84 (0.13)</td>
<td>5.96 (0.73)</td>
</tr>
<tr>
<td>Kernel</td>
<td>LCV</td>
<td>1.91 (0.09)</td>
<td>1.70 (0.11)</td>
<td>5.03 (0.73)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LSCV</td>
<td>1.88 (0.11)</td>
<td>1.40 (0.10)</td>
<td>3.96 (0.38)</td>
</tr>
<tr>
<td>Mixture</td>
<td>LCV</td>
<td>1.56 (0.09)</td>
<td>1.28 (0.10)</td>
<td>3.49 (0.32)</td>
</tr>
</tbody>
</table>

**NOTE:** The values in parentheses are the corresponding standard errors.
Part V
A methodology for a two-component mixture model with symmetrized nonparametric components is proposed.

Simulation results and real examples show that the mixture-based methods are more appealing than the kernel-based methods.

With advantages such as:

- greater flexibility
- simpler final model
- better performance

the mixture-based methods are competitive for practical applications.
Acknowledgements

- MOHE and UMT – sponsoring my study
- Dept. of Stats., The UoA – covering conference-related costs
References

**Cook, R. D. and Weisberg, S.**
An Introduction to Regression Graphics

**Burnham, K. P. and Anderson, D. R.**
Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition
*New York: Springer-Verlag, 2002.*

**Heinz, G., Peterson, L. J., Johnson, R. W. and Kerk, C. J.**
Exploring Relationships in Body Dimensions

**Bordes, L., Mottelet, S. and Vandekerkhove, P.**
Semiparametric Estimation of a Two-Component Mixture Model

**Hunter, D. R., Wang, S. and Hettmansperger, T. P.**
Inference for Mixtures of Symmetric Distributions
Bordes, L., Chauveau, D. and Vandekerkhove, P.
A stochastic EM algorithm for a semiparametric mixture model

Benaglia, T., Chauveau, D. and Hunter, D. R.
An EM-Like Algorithm for Semi- and Nonparametric Estimation in Multivariate Mixtures

Wang, Y.
Maximum Likelihood Computation for Fitting Semiparametric Mixture Models

Yee, T. W.
VGLMs and VGAMs: an Overview for Applications in Fisheries Research
*Fish. Res.*, 2009, Accepted.