Bandwidth selectors for multivariate kernel density estimation

Le choix de fenêtres pour l'estimation de la densité multivariée à noyau

Tarn Duong
Supervisor/Directeur: Dr Martin Hazelton
University of Western Australia

August/Août 2004

Outline

1. Motivation (finite sample)

2. Optimal bandwidth selectors

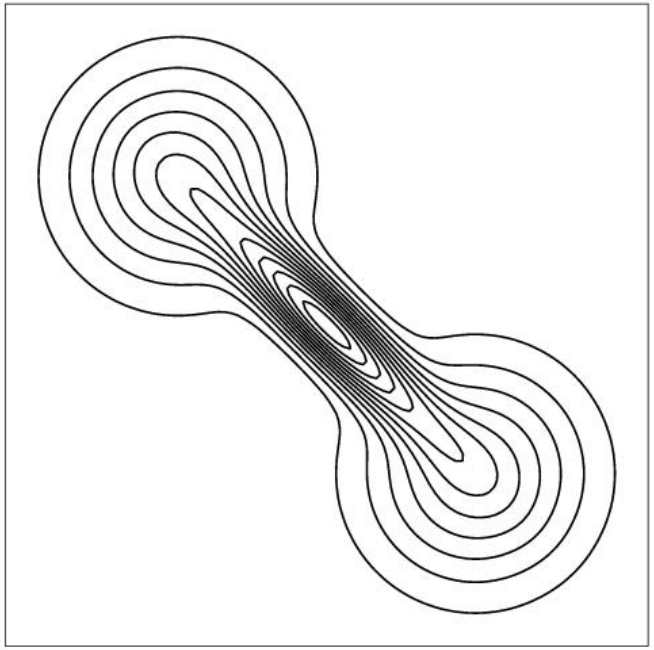
3. Asymptotic behaviour

4. Application to real data

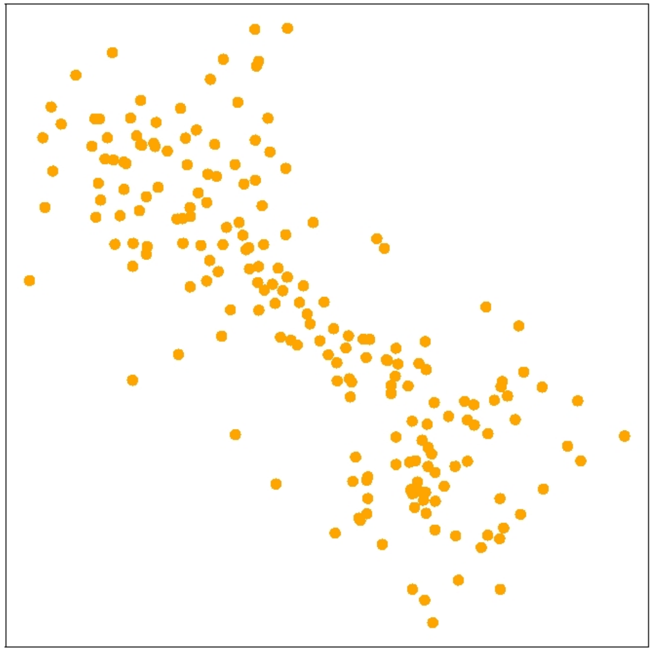
5. Summary

Motivation – dumbbell density

Contour plot



Sample



Properties of KDE

- non-parametric
- easy to compute
- easy to interpret

Equation for KDE

Kernel density estimate (KDE) is

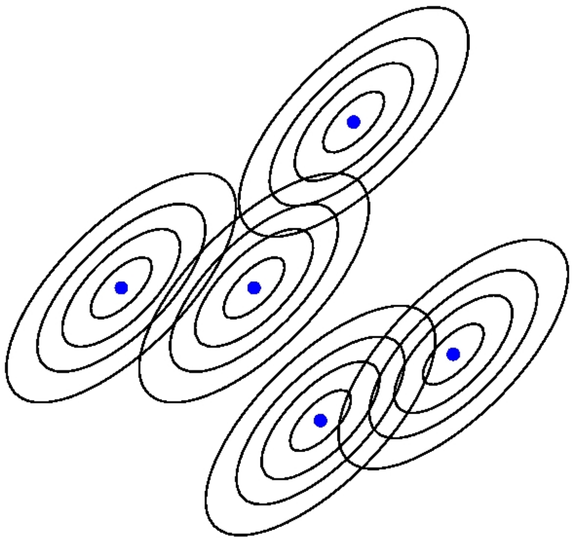
$$\hat{f}(\boldsymbol{x}; \mathbf{H}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}(\boldsymbol{x} - \boldsymbol{X}_i)$$

where

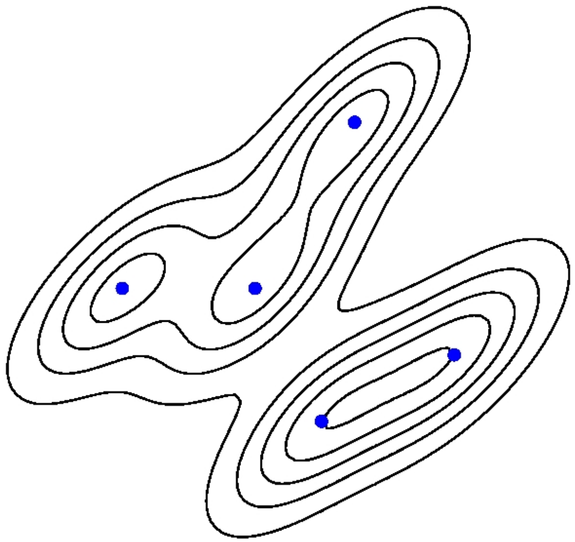
- $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is random sample of n d-variate data points
- $\mathbf{H}$ is **bandwidth** matrix parameter
- $K_{\mathbf{H}}(\cdot)$ is normal pdf with mean $\mathbf{0}$ and variance $\mathbf{H}$

Constructing KDE

Individual kernels



Averaged kernels = KDE

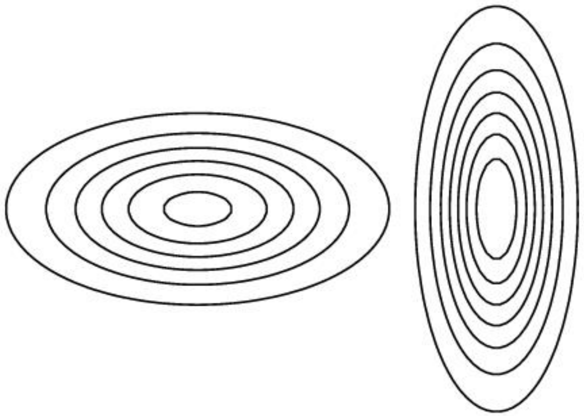


Bandwidth selection

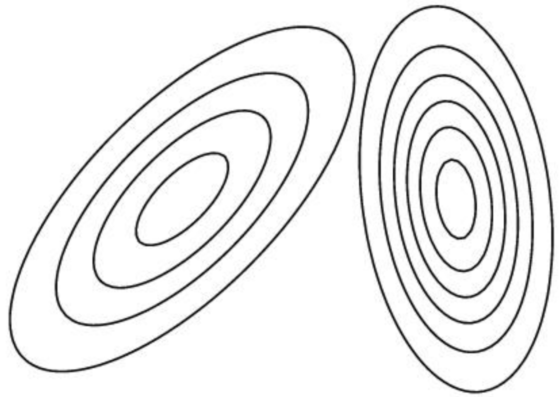
- single most important factor affecting performance of KDE
- induces orientation of kernel
- controls spread of kernel
- diagonal bandwidth $\begin{bmatrix} h_1^2 & 0 \\ 0 & h_2^2 \end{bmatrix}$ or full bandwidth $\begin{bmatrix} h_1^2 & h_{12} \\ h_{12} & h_2^2 \end{bmatrix}$

Kernel orientation

Diagonal bandwidth matrix

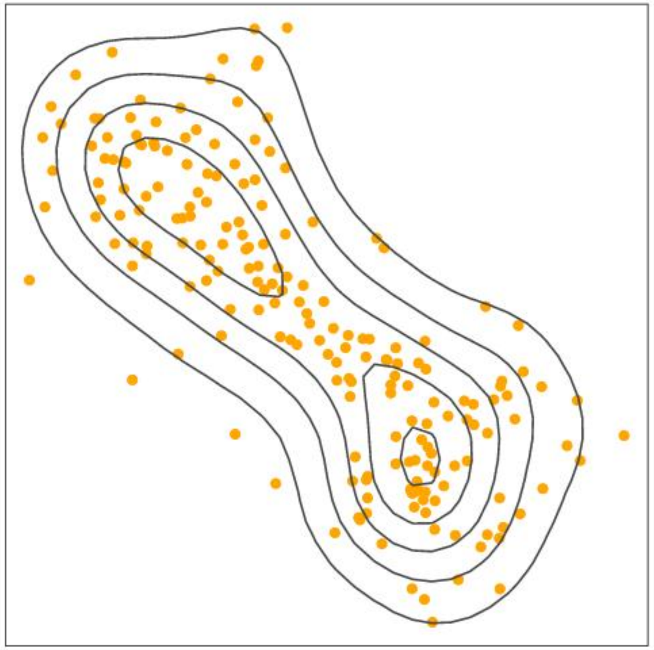


Full bandwidth matrix

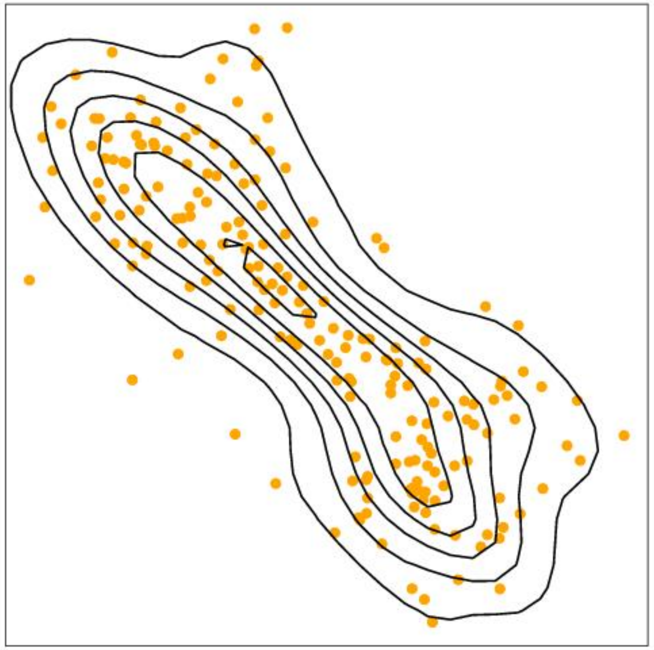


KDE of simulated dumbbell density

Diagonal bandwidth matrix



Full bandwidth matrix



Error measures (1)

Mean Integrated Squared Error (MISE) is

$$\mathrm{MISE}(\mathbf{H}) = \int_{\mathbb{R}^d} \mathbb{E}[\hat{f}(\boldsymbol{x};\mathbf{H}) - f(\boldsymbol{x})]^2 \, d\boldsymbol{x}$$

where

- $\hat{f}(\boldsymbol{x};\mathbf{H})$ is kernel density estimate
- $f(\boldsymbol{x})$ is (unknown) target density

Error measures (2)

Asymptotic Mean Integrated Squared Error (AMISE) is

$$\mathrm{AMISE}(\mathbf{H}) = \underbrace{n^{-1}R(K)|\mathbf{H}|^{-1/2}}_{\text{Asymptotic integrated variance}} + \underbrace{\int_{\mathbb{R}^d} [(K_{\mathbf{H}} * f)(\boldsymbol{x}) - f(\boldsymbol{x})]^2 \, d\boldsymbol{x}}_{\text{Exact integrated squared bias}}$$

where $R(K) = \int_{\mathbb{R}^d} K(\boldsymbol{x})^2 \, d\boldsymbol{x}$ and $*$ is the convolution operator.

Optimal bandwidth selectors

- MISE-optimal bandwidth selector is

$$\mathbf{H}_{\text{MISE}} = \underset{\mathbf{H}}{\text{argmin}} \; \text{MISE}(\mathbf{H}).$$

- Our surrogate for this is

$$\mathbf{H}_{\text{AMISE}} = \underset{\mathbf{H}}{\text{argmin}} \; \text{AMISE}(\mathbf{H}).$$

- Our data-driven bandwidth selector is

$$\hat{\mathbf{H}} = \underset{\mathbf{H}}{\text{argmin}} \; \widehat{\text{AMISE}}(\mathbf{H}).$$

Smoothed cross validation (1)

Obtain SCV estimate of AMISE by replacing f by a pilot kernel density estimate $\hat{f}_P$

$$\hat{f}_P(\boldsymbol{x}; \mathbf{G}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{G}}(\boldsymbol{x} - \boldsymbol{X}_i)$$

where $\mathbf{G}$ is pilot bandwidth matrix.

- same as $\hat{f}$ except for $\mathbf{G}$
- can be quite imprecise since is only a preliminary estimate
- $\hat{f}_P$ is refined to more accurate $\hat{f}$

Smoothed cross validation (2)

$$\mathrm{SCV}(\mathbf{H};\mathbf{G}) = n^{-1}R(K)|\mathbf{H}|^{-1/2} + n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(K_{2\mathbf{H}+2\mathbf{G}} - 2K_{\mathbf{H}+2\mathbf{G}} + K_{2\mathbf{G}})(\boldsymbol{X}_i - \boldsymbol{X}_j)$$

where K is normal kernel.

SCV algorithms

- Hall, Marron & Park (1992)'s univariate selector, with optimal pilot selector g independent of h
- Sain, Baggerly & Scott (1994)'s multivariate selector, diagonal matrix with sub-optimal pilot selector set equal to $\mathbf{H}$
- Proposed: multivariate selector, full matrix with optimal pilot selector $g^2\mathbf{I}$ independent of $\mathbf{H}$

Pilot bandwidth matrix selection (1)

- Simplify problem first by using $\mathbf{G} = g^2\mathbf{I}$
- Appropriate value of g

$$g_0 = \underset{g>0}{\operatorname{argmin}} \ \operatorname{MSE}(\hat{\mathbf{H}}(g))$$

where

$$\begin{aligned}\operatorname{MSE}(\hat{\mathbf{H}}(g)) = g^4 C_1 + 2n^{-1} g^{-d-2} C_2 + n^{-2} g^{-2d-8} C_3 \\ + O(n^{-2} g^{-2d-8} + n^{-1})\end{aligned}$$

Pilot bandwidth matrix selection (2)

- $\mathrm{MSE}(\hat{\mathbf{H}}(g))$ has minimum at

$$g_0 = \left[\frac{2(d+4)C_3}{n\sqrt{-(d+2)C_2 + C_1}}\right]^{1/(d+6)}$$

where C_1, C_2 and C_3 are constants that involve d, f, K but not n.

- Estimate C_1, C_2, C_3 and 'plug-in' estimates to give $\hat{g}_0$.

Proposed SCV algorithm

1. Set pilot bandwidth to $\hat{g}_0^2\mathbf{I}$.

2. Form $\mathrm{SCV}(\mathbf{H}; \hat{g}_0^2\mathbf{I})$

3. $\hat{\mathbf{H}} = \operatorname*{argmin}_{\mathbf{H}} \mathrm{SCV}(\mathbf{H}; \hat{g}_0^2\mathbf{I})$

Order in probability – univariate

Let $\{a_n\}, \{b_n\}$ be sequence of numbers. Then

$$a_n = O(b_n) \text{ if } \exists M, L \,:\, n > M \Rightarrow |a_n| < L|b_n|$$

Let $\{A_n\}$ be sequence of random variables. Then

$$A_n = O_p(b_n) \text{ if } \forall \epsilon > 0, \exists M, L \,:\, n > M \Rightarrow \mathbb{P}(|A_n| < L|b_n|) > 1 - \epsilon$$

Relative convergence rate – univariate

$\hat{h}$ converges to h_{AMISE} with (relative) rate $n^{-\alpha}$ if

$$\hat{h} - h_{\text{AMISE}} = O_p(n^{-\alpha})h_{\text{AMISE}}$$

$\hat{h}$ converges to h_{MISE} with (relative) rate $n^{-\alpha}$ if

$$\hat{h} - h_{\text{MISE}} = O_p(n^{-\alpha})h_{\text{MISE}}$$

Order in probability – multivariate

Let $\{\mathbf{A}_n\}, \{\mathbf{B}_n\}$ be sequences of matrices of the same dimensions. Then

$$\mathbf{A}_n = O_p(\mathbf{B}_n) \text{ if } a_{n,ij} = O_p(b_{n,ij})$$

for all elements $a_{n,ij}$ of $\mathbf{A}_n$ and $b_{n,ij}$ of $\mathbf{B}_n$.

Relative convergence rate – multivariate

- $\hat{\mathbf{H}}$ converges to $\mathbf{H}_{\text{AMISE}}$ with (relative) rate $n^{-\alpha}$ if

$$\text{vech}(\hat{\mathbf{H}} - \mathbf{H}_{\text{AMISE}}) = O_p(n^{-\alpha}\mathbf{J})\,\text{vech}\,\mathbf{H}_{\text{AMISE}}$$

  where

  – vech is vector half operator e.g. if $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\text{vech}\,\mathbf{A} = \begin{bmatrix} a & c & d \end{bmatrix}^T$

  – $\mathbf{J}$ is $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix of ones

- Corresponding definition for $\mathbf{H}_{\text{MISE}}$.

Strategy (1)

Let $\text{MSE}(\hat{\mathbf{H}}) = \mathbb{E}[\text{vech}(\hat{\mathbf{H}} - \mathbf{H}_{\text{AMISE}}) \, \text{vech}^T(\hat{\mathbf{H}} - \mathbf{H}_{\text{AMISE}})]$.

If

$$\text{MSE}(\hat{\mathbf{H}}) = O(n^{-2\alpha}\mathbf{J})(\text{vech}\, \mathbf{H}_{\text{AMISE}})(\text{vech}^T \, \mathbf{H}_{\text{AMISE}})$$

then $\hat{\mathbf{H}}$ has relative rate $n^{-\alpha}$ of convergence to $\mathbf{H}_{\text{AMISE}}$.

Strategy (2)

We have

$$\mathrm{MSE}(\hat{\mathbf{H}}) = \mathrm{Var}(\hat{\mathbf{H}}) + [\mathrm{Bias}(\hat{\mathbf{H}})][\mathrm{Bias}^T(\hat{\mathbf{H}})]$$

where

$$\mathrm{Bias}(\hat{\mathbf{H}}) = O\left(\mathbb{E}\left[\frac{\partial}{\partial\, \mathrm{vech}\, \mathbf{H}} (\widehat{\mathrm{AMISE}} - \mathrm{AMISE})(\mathbf{H}_{\mathrm{AMISE}}) \right] \right)$$

$$\mathrm{Var}(\hat{\mathbf{H}}) = O\left(\mathrm{Var}\left[\frac{\partial}{\partial\, \mathrm{vech}\, \mathbf{H}} (\widehat{\mathrm{AMISE}} - \mathrm{AMISE})(\mathbf{H}_{\mathrm{AMISE}}) \right] \right)$$

Strategy (3)

1. Compute order of expected value and variance of

$$\frac{\partial}{\partial \,\mathrm{vech}\, \mathbf{H}}(\widehat{\mathrm{AMISE}} - \mathrm{AMISE})(\mathbf{H}_{\mathrm{AMISE}}).$$

2. Compute $\mathrm{MSE}(\hat{\mathbf{H}})$ from Step 1. Convergence rate to $\mathbf{H}_{\mathrm{AMISE}}$ follows immediately.

3. If rate of $\mathbf{H}_{\mathrm{AMISE}}$ to $\mathbf{H}_{\mathrm{MISE}} <$ rate of $\hat{\mathbf{H}}$ to $\mathbf{H}_{\mathrm{AMISE}}$ then convergence rate of $\hat{\mathbf{H}}$ to $\mathbf{H}_{\mathrm{MISE}}$ is same as in Step 2.

Convergence rate for SCV selector (1)

Step 1. (a)

$$\mathbb{E}\left[\frac{\partial}{\partial \operatorname{vech} \mathbf{H}}(\mathrm{SCV} - \mathrm{AMISE})(\mathbf{H}_{\mathrm{AMISE}})\right]$$
$$= O((g_0^2 + n^{-1}g_0^{-d-4})\mathbf{J}) \operatorname{vech} \mathbf{H}_{\mathrm{AMISE}}$$

as

$$\mathbb{E}[\mathrm{SCV}(\mathbf{H}; g^2\mathbf{I})] = \mathrm{AMISE}(\mathbf{H}) + O((g^2 + n^{-1}g^{-d-4})\|\operatorname{vech} \mathbf{H}\|^2)$$

Convergence rate for SCV selector (2)

Step 1. (b)

$$\mathrm{Var}\left[\frac{\partial}{\partial\,\mathrm{vech}\,\mathbf{H}}(\mathrm{SCV}-\mathrm{AMISE})(\mathbf{H}_{\mathrm{AMISE}})\right]$$
$$= O((n^{-2}g_0^{-d-8}+n^{-1})\mathbf{J})(\mathrm{vech}\,\mathbf{H}_{\mathrm{AMISE}})(\mathrm{vech}^T\,\mathbf{H}_{\mathrm{AMISE}})$$

as

$$\mathrm{Var}[\dots] = n^{-2}V_1 + n^{-1}V_2$$

where

$$V_1 = O(g^{-d-8}\mathbf{J})(\mathrm{vech}\,\mathbf{H})(\mathrm{vech}^T\,\mathbf{H})$$
$$V_2 = O(\mathbf{J})(\mathrm{vech}\,\mathbf{H})(\mathrm{vech}^T\,\mathbf{H})$$

Convergence rate for SCV selector (3)

Step 1. (cont.)

$$\mathrm{Bias}(\hat{\mathbf{H}}) = O(n^{-2/(d+6)}\mathbf{J})\,\mathrm{vech}\,\mathbf{H}_{\mathrm{AMISE}}$$

$$\mathrm{Var}(\hat{\mathbf{H}}) = O(n^{-4/(d+6)}\mathbf{J})(\mathrm{vech}\,\mathbf{H}_{\mathrm{AMISE}})(\mathrm{vech}^T\,\mathbf{H}_{\mathrm{AMISE}})$$

Step 2.

$$\mathrm{MSE}(\hat{\mathbf{H}}) = O(n^{-4/(d+6)}\mathbf{J})(\mathrm{vech}\,\mathbf{H}_{\mathrm{AMISE}})(\mathrm{vech}^T\,\mathbf{H}_{\mathrm{AMISE}})$$

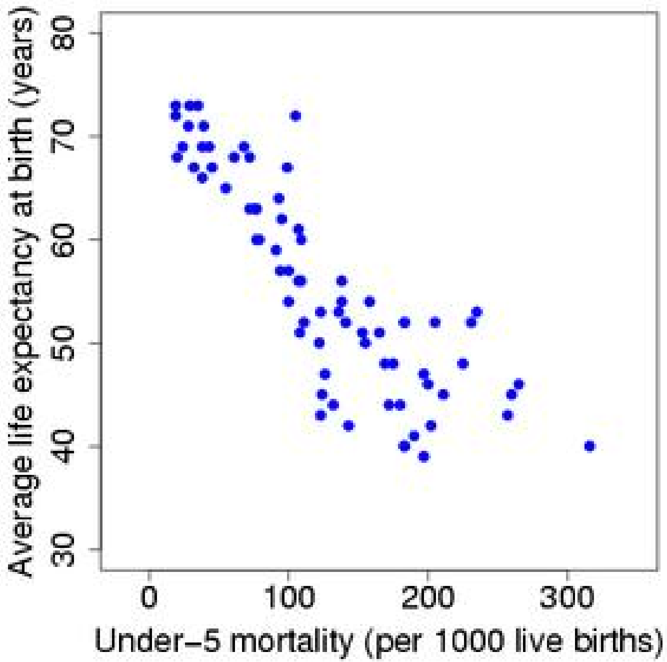
$\Rightarrow \hat{\mathbf{H}}$ has convergence rate $n^{-2/(d+6)}$ to $\mathbf{H}_{\mathrm{AMISE}}$.

Convergence rate for SCV selector (4)

Step 3.

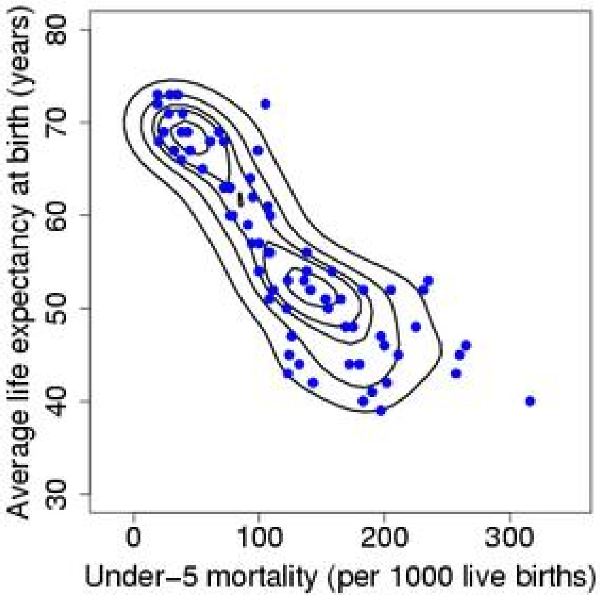
- Rate of $\mathbf{H}_{\text{AMISE}}$ to $\mathbf{H}_{\text{MISE}}$ is $n^{-2/(d+4)}$.
- This is smaller than $n^{-2/(d+6)}$ for all d.
- So $\hat{\mathbf{H}}$ has rate $n^{-2/(d+6)}$ to $\mathbf{H}_{\text{MISE}}$ for all d.

Unicef data (1)

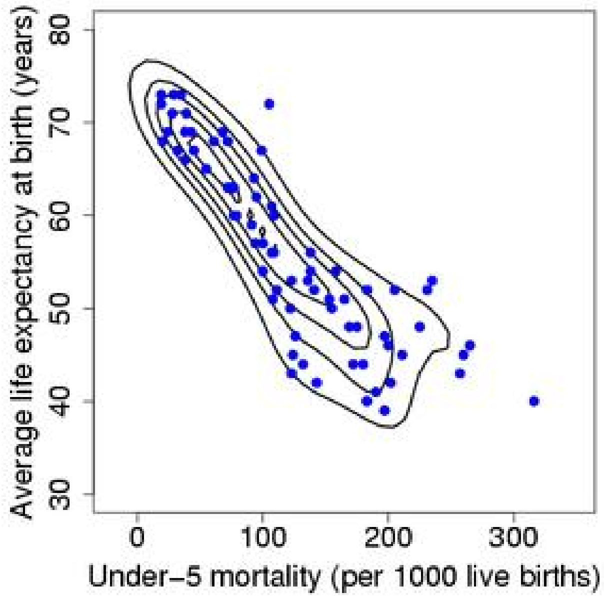


Unicef data (2)

Diagonal selector



Full SCV selector



Summary and future directions

Summary

- Constructed full SCV selector (includes optimal scalar pilot selector)
- Supplied asymptotic relative convergence rate
- Shown good finite sample behaviour with simulated and real data

Future directions

- Develop full matrix pilot selector instead of current scalar selector