Parametric or Nonparametric: The Focused Information Criterion Approach

Martin Jullum

Nils Lid Hjort

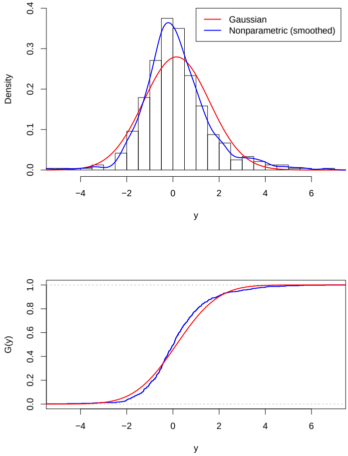
University of Oslo

martinju@math.uio.no

August 4, 2014

Modeling and model selection

- Unknown underlying distribution G for data $Y_1, \ldots, Y_n$
- Parametric approach
- Nonparametric approach
- Several appropriate models – which one should we trust?



Classical model selection

- Information criteria
  - AIC, BIC, DIC, GIC, TIC,. . .
  - Cannot handle nonparametrics
- Goodness-of-fit measures
  - Cramér–von Mises, Kolmogorov–Smirnov, Pearson's chi-squared
  - Requires a pre-set significance level

Classical model selection

- Information criteria
  - AIC, BIC, DIC, GIC, TIC,. . .
  - Cannot handle nonparametrics
- Goodness-of-fit measures
  - Cramér–von Mises, Kolmogorov–Smirnov, Pearson's chi-squared
  - Requires a pre-set significance level

Claim and objective

- No particularly well-suited criterion for model selection among parametrics and nonparametrics
- Different models have strengths and weaknesses at different parts of the data space
- What you want to learn should reflect your choice of model

Research objective

Create a focused/interest driven model selection criterion for selection among a set of parametric and nonparametric models

Claim and objective

- No particularly well-suited criterion for model selection among parametrics and nonparametrics
- Different models have strengths and weaknesses at different parts of the data space
- What you want to learn should reflect your choice of model

Research objective

Create a focused/interest driven model selection criterion for selection among a set of parametric and nonparametric models

Claim and objective

- No particularly well-suited criterion for model selection among parametrics and nonparametrics
- Different models have strengths and weaknesses at different parts of the data space
- What you want to learn should reflect your choice of model

Research objective

Create a focused/interest driven model selection criterion for selection among a set of parametric and nonparametric models

Focus parameter

- A general population quantity of interest – not(!) a model specific parameter
- A functional μ of the distribution G: $\mu(G)$

Examples

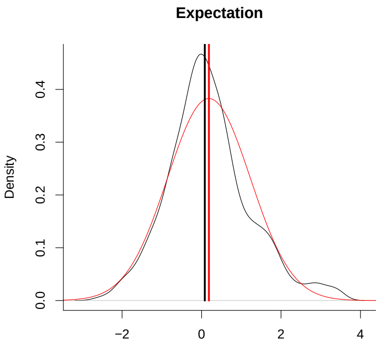
- Expectation: $\mu(G) = \mathrm{E}_G[Y_i]$
- $Pr\{Y_i > 2\}$: $\mu(G) = 1 - G(2)$
- Interquartile range: $\mu(G) = G^{-1}(3/4) - G^{-1}(1/4)$

Focus parameter

- A general population quantity of interest – not(!) a model specific parameter
- A functional μ of the distribution G: $\mu(G)$

Examples

- Expectation: $\mu(G) = \mathrm{E}_G[Y_i]$
- $Pr\{Y_i > 2\}$: $\mu(G) = 1 - G(2)$
- Interquartile range: $\mu(G) = G^{-1}(3/4) - G^{-1}(1/4)$

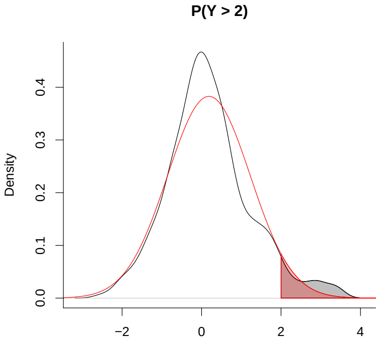


Focus parameter

- A general population quantity of interest – not(!) a model specific parameter
- A functional μ of the distribution G: $\mu(G)$

Examples

- Expectation: $\mu(G) = \mathrm{E}_G[Y_i]$
- $Pr\{Y_i > 2\}$: $\mu(G) = 1 - G(2)$
- Interquartile range: $\mu(G) = G^{-1}(3/4) - G^{-1}(1/4)$

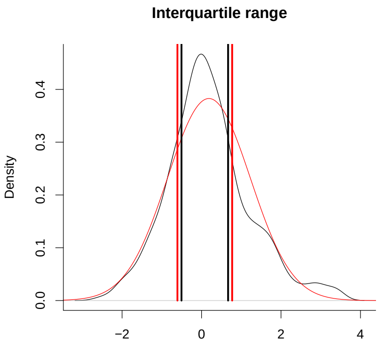


Focus parameter

- A general population quantity of interest – not(!) a model specific parameter
- A functional μ of the distribution G: $\mu(G)$

Examples

- Expectation: $\mu(G) = \mathrm{E}_G\left[Y_i\right]$
- $Pr\left\{Y_i > 2\right\}$: $\mu(G) = 1 - G(2)$
- Interquartile range: $\mu(G) = G^{-1}(3/4) - G^{-1}(1/4)$



Criterion idea

- Model selection problem → Best estimator for $\mu = \mu(G)$
- Estimate μ by plug-in estimation for each model M: $\widehat{\mu}_M = \mu(\widehat{G}_M)$
- Performance measure:

$$\mathrm{mse}(\widehat{\mu}_M) = \mathsf{E}\left[(\widehat{\mu}_M - \mu_{\mathrm{true}})^2\right] = \mathrm{bias}^2(\widehat{\mu}_M) + \mathrm{Var}(\widehat{\mu}_M)$$

Basic idea

- Estimate the mean squared error (mse) as squared bias + variance:

$$\mathsf{FIC}(M) = \widehat{\mathrm{mse}}(\widehat{\mu}_M) = \widehat{\mathrm{bias}^2}(\widehat{\mu}_M) + \widehat{\mathrm{Var}}(\widehat{\mu}_M)$$

- Choose the model and estimator with the smallest estimated mse

- Conceptually similar approach as the original FIC criterion by Claeskens and Hjort (JASA 2003)

Criterion idea

- Model selection problem $\rightarrow$ Best estimator for $\mu = \mu(G)$
- Estimate μ by plug-in estimation for each model M: $\widehat{\mu}_M = \mu(\widehat{G}_M)$
- Performance measure:

$$\mathrm{mse}(\widehat{\mu}_M) = \mathsf{E}\left[(\widehat{\mu}_M - \mu_{\mathrm{true}})^2\right] = \mathrm{bias}^2(\widehat{\mu}_M) + \mathrm{Var}(\widehat{\mu}_M)$$

Basic idea

- Estimate the mean squared error (mse) as squared bias + variance:

$$\mathsf{FIC}(M) = \widehat{\mathrm{mse}}(\widehat{\mu}_M) = \widehat{\mathrm{bias}^2}(\widehat{\mu}_M) + \widehat{\mathrm{Var}}(\widehat{\mu}_M)$$

- Choose the model and estimator with the smallest estimated mse

- Conceptually similar approach as the original FIC criterion by Claeskens and Hjort (JASA 2003)

Mse estimation

- Mse estimation based on joint limiting distribution for focus parameter estimators
- Use straight forward empirical analogues of unknown quantities
- Precise formulae depend on the type of data situation
- Data types
  - Univariate and multivariate i.i.d. data, categorical data, time series
  - Data from different populations
  - Hazard rate models
  - Cox regression vs. parametric regression
- Other strategies for the standard regression setting and density estimation

Mse estimation

- Mse estimation based on joint limiting distribution for focus parameter estimators
- Use straight forward empirical analogues of unknown quantities
- Precise formulae depend on the type of data situation
- Data types
  - Univariate and multivariate i.i.d. data, categorical data, time series
  - Data from different populations
  - Hazard rate models
  - Cox regression vs. parametric regression
- Other strategies for the standard regression setting and density estimation

Mse estimation

- Mse estimation based on joint limiting distribution for focus parameter estimators
- Use straight forward empirical analogues of unknown quantities
- Precise formulae depend on the type of data situation
- Data types
  - Univariate and multivariate i.i.d. data, categorical data, time series
  - Data from different populations
  - Hazard rate models
  - Cox regression vs. parametric regression
- Other strategies for the standard regression setting and density estimation

Mse estimation

- Mse estimation based on joint limiting distribution for focus parameter estimators
- Use straight forward empirical analogues of unknown quantities
- Precise formulae depend on the type of data situation
- Data types
  - Univariate and multivariate i.i.d. data, categorical data, time series
  - Data from different populations
  - Hazard rate models
  - Cox regression vs. parametric regression
- Other strategies for the standard regression setting and density estimation

Averaged Focused Information Criterion (AFIC)

- Sometimes a single focus parameter is not desirable
- Focus on a (weighted) set of focus parameters simultaneously
- Performance measure:
  risk $= \int \mathrm{mse}(\widehat{\mu}(t))\, \mathrm{d}W(t) = \int \mathrm{E}\left[(\widehat{\mu}(t) - \mu_{\mathsf{true}}(t))^2\right] \mathrm{d}W(t)$,

General AFIC formula

$$\mathsf{AFIC}(M) = \int \widehat{\mathrm{mse}}(\widehat{\mu}_M(t))\, \mathrm{d}W(t) = \int \mathrm{FIC}(\widehat{\mu}_M(t))\, \mathrm{d}W(t)$$

Averaged Focused Information Criterion (AFIC)

- Sometimes a single focus parameter is not desirable
- Focus on a (weighted) set of focus parameters simultaneously
- Performance measure:
  risk $= \int \mathrm{mse}(\widehat{\mu}(t))\,\mathrm{d}W(t) = \int \mathsf{E}\left[(\widehat{\mu}(t) - \mu_{\mathsf{true}}(t))^2\right]\mathrm{d}W(t)$,

General AFIC formula

$$\mathsf{AFIC}(M) = \int \widehat{\mathrm{mse}}(\widehat{\mu}_M(t))\,\mathrm{d}W(t) = \int \mathrm{FIC}(\widehat{\mu}_M(t))\,\mathrm{d}W(t)$$

Criterion properties (i.i.d.)

Robust mse estimation

- Consistent variance and squared bias estimators
- Estimation outside model conditions

FIC asymptotics

- Parametrics biased: $Pr\left\{\text{Select pm}\right\} \to 0$
- Parametrics correct: $Pr\left\{\text{Select pm}\right\} \to \chi_1^2(2) \approx 0.843$

AFIC asymptotics are not as general and depend on μ and W

Criterion properties (i.i.d.)

Robust mse estimation

- Consistent variance and squared bias estimators
- Estimation outside model conditions

FIC asymptotics

- Parametrics biased: $Pr\,\{\text{Select pm}\} \to 0$
- Parametrics correct: $Pr\,\{\text{Select pm}\} \to \chi^2_1(2) \approx 0.843$

AFIC asymptotics are not as general and depend on μ and W

Criterion properties (i.i.d.)

Robust mse estimation

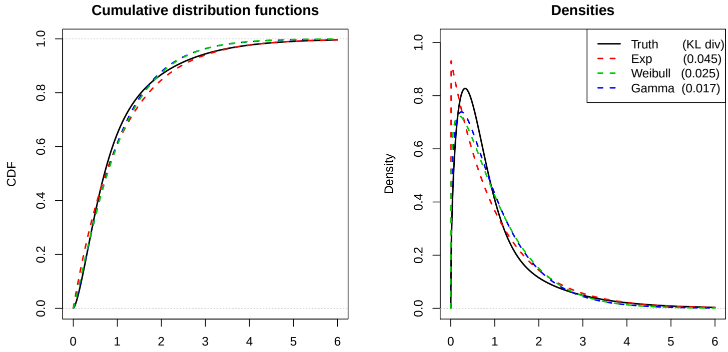
- Consistent variance and squared bias estimators
- Estimation outside model conditions

FIC asymptotics

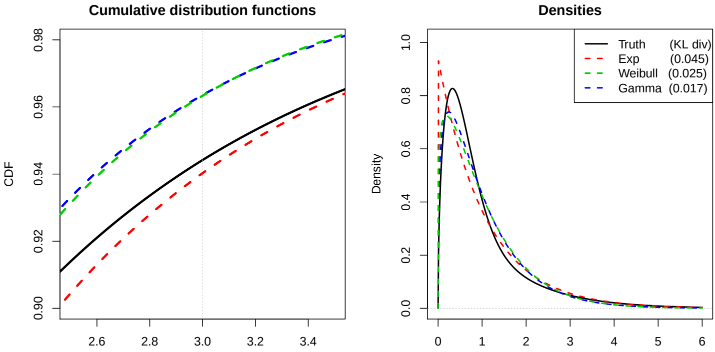
- Parametrics biased: $Pr\,\{\text{Select pm}\} \to 0$
- Parametrics correct: $Pr\,\{\text{Select pm}\} \to \chi_1^2(2) \approx 0.843$

AFIC asymptotics are not as general and depend on μ and W

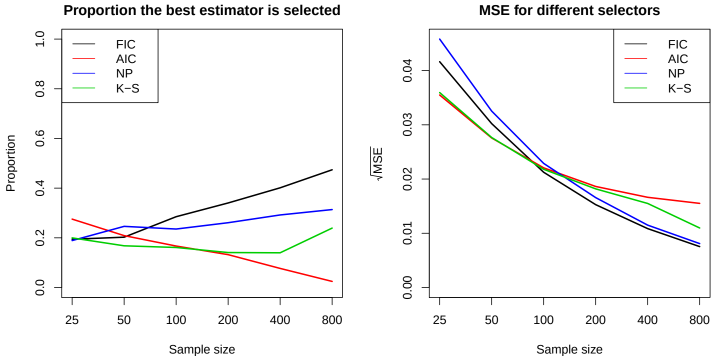
Performance comparison



Performance comparison



Performance comparison



Summary

- Focus driven model selection with a nonparametric alternative
- Rank models according to $\widehat{\mu}_M$'s estimated risk
- Robustifies parametric model selection by including the nonparametric candidate model
- AFIC allows several focus parameters to be considered simultaneously
- Optimistic performance comparison results

R-functions: folk.uio.no/martinju/FIC

Contact: martinju@math.uio.no