FIC with a nonparametric candidate
– a new strategy for FIC construction –

Martin Jullum

Norwegian Computing Center

jullum@nr.no

May 9, 2016

The general FIC idea

Classical model selection

- AIC, BIC, DIC,...
- Overall measuring information criterion

Focused Information Criterion (FIC)

- What you want to learn from the data analysis should reflect your choice of model
- Assumes the purpose of the analysis is to estimate a specific focus parameter μ
- Model selection problem $\rightarrow$ Best estimator for μ
- Performance measure: Risk of μ-estimator (typically MSE-type)
- The FIC score is an estimate of the this risk
- Models/estimators are ranked by their FIC scores

The general FIC idea

Classical model selection

- AIC, BIC, DIC,...
- Overall measuring information criterion

Focused Information Criterion (FIC)

- What you want to learn from the data analysis should reflect your choice of model
- Assumes the purpose of the analysis is to estimate a specific focus parameter μ
- Model selection problem $\rightarrow$ Best estimator for μ
- Performance measure: Risk of μ-estimator (typically MSE-type)
- The FIC score is an estimate of the this risk
- Models/estimators are ranked by their FIC scores

The original FIC (I)

- Assume all candidate models are parametrically nested (i.i.d. data)
    - Wide model (θ: free, γ: free), density: $f(y;\theta,\gamma)$, focus: $\widehat{\mu}_{\text{wide}} = \mu(\widehat{\theta}_{\text{wide}}, \widehat{\gamma}_{\text{wide}})$
    - Narrow model (θ: free, γ: fixed $= \gamma_0$), density: $f(y;\theta,\gamma_0)$, focus: $\widehat{\mu}_{\text{narr}} = \mu(\widehat{\theta}_{\text{narr}}, \gamma_0)$
    - All other models assumed to lie between these, with different portions of the full γ parameter estimated/fixed
- Relies on a local misspecification framework: $g_n(y) = f(y;\theta_0,\gamma_0+\delta/\sqrt{n})$
- True value of μ is $\mu_{\text{true},n} = \mu(\theta_0,\gamma_0+\delta/\sqrt{n})$
- Under this framework, for each submodel S:

$$\Lambda_{S,n} = \sqrt{n}(\widehat{\mu}_S - \mu_{\text{true},n}) \to_d \Lambda_S \sim \mathrm{N}(\text{bias}_S(\delta), \text{var}_S(\delta))$$

The original FIC (I)

- Assume all candidate models are parametrically nested (i.i.d. data)
  - Wide model (θ: free, γ: free), density: $f(y;\theta,\gamma)$, focus: $\widehat{\mu}_{\text{wide}} = \mu(\widehat{\theta}_{\text{wide}}, \widehat{\gamma}_{\text{wide}})$
  - Narrow model (θ: free, γ: fixed $= \gamma_0$), density: $f(y;\theta,\gamma_0)$, focus: $\widehat{\mu}_{\text{narr}} = \mu(\widehat{\theta}_{\text{narr}}, \gamma_0)$
  - All other models assumed to lie between these, with different portions of the full γ parameter estimated/fixed
- Relies on a local misspecification framework: $g_n(y) = f(y;\theta_0,\gamma_0+\delta/\sqrt{n})$
- True value of μ is $\mu_{\mathrm{true},n} = \mu(\theta_0,\gamma_0+\delta/\sqrt{n})$
- Under this framework, for each submodel S:

$$\Lambda_{S,n} = \sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n}) \to_d \Lambda_S \sim \mathrm{N}(\text{bias}_S(\delta), \text{var}_S(\delta))$$

The original FIC (I)

- Assume all candidate models are parametrically nested (i.i.d. data)
  - Wide model (θ: free, γ: free), density: $f(y;\theta,\gamma)$, focus: $\widehat{\mu}_{\text{wide}} = \mu(\widehat{\theta}_{\text{wide}}, \widehat{\gamma}_{\text{wide}})$
  - Narrow model (θ: free, γ: fixed $= \gamma_0$), density: $f(y;\theta,\gamma_0)$, focus: $\widehat{\mu}_{\text{narr}} = \mu(\widehat{\theta}_{\text{narr}}, \gamma_0)$
  - All other models assumed to lie between these, with different portions of the full γ parameter estimated/fixed
- Relies on a local misspecification framework: $g_n(y) = f(y;\theta_0,\gamma_0+\delta/\sqrt{n})$
- True value of μ is $\mu_{\mathrm{true},n} = \mu(\theta_0,\gamma_0+\delta/\sqrt{n})$
- Under this framework, for each submodel S:

$$\Lambda_{S,n} = \sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n}) \to_d \Lambda_S \sim \mathrm{N}(\text{bias}_S(\delta), \text{var}_S(\delta))$$

The original FIC (II)

- Risk: MSE of the limiting distribution of $\sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n})$ under the local misspecification framework, i.e. for submodel S:

$$\mathsf{risk}_S = \mathsf{mse}(\Lambda_S) = \mathsf{bias}_S^2(\delta) + \mathsf{var}_S(\delta)$$

- Estimates the risk

$$\mathsf{FIC}_S = \widehat{\mathsf{risk}}_S = \widehat{\mathsf{bias}_S^2(\delta)} + \widehat{\mathsf{var}_S(\delta)}$$

- Constructed for
  - I.i.d. and standard regression models
  - Variable selection for Cox's prop. haz. Aalen's lin. haz. regression (with censoring)
  - Parametric time series models
  - etc.

The original FIC (II)

- Risk: MSE of the limiting distribution of $\sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n})$ under the local misspecification framework, i.e. for submodel S:

$$\mathsf{risk}_S = \mathsf{mse}(\Lambda_S) = \mathsf{bias}_S^2(\delta) + \mathsf{var}_S(\delta)$$

- Estimates the risk

$$\mathsf{FIC}_S = \widehat{\mathsf{risk}}_S = \widehat{\mathsf{bias}_S^2(\delta)} + \widehat{\mathsf{var}_S(\delta)}$$

- Constructed for
  - I.i.d. and standard regression models
  - Variable selection for Cox's prop. haz. Aalen's lin. haz. regression (with censoring)
  - Parametric time series models
  - etc.

The original FIC (II)

- Risk: MSE of the limiting distribution of $\sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n})$ under the local misspecification framework, i.e. for submodel S:

$$\mathsf{risk}_S = \mathsf{mse}(\Lambda_S) = \mathsf{bias}_S^2(\delta) + \mathsf{var}_S(\delta)$$

- Estimates the risk

$$\mathsf{FIC}_S = \widehat{\mathsf{risk}}_S = \widehat{\mathsf{bias}_S^2(\delta)} + \widehat{\mathsf{var}_S(\delta)}$$

- Constructed for
  - I.i.d. and standard regression models
  - Variable selection for Cox's prop. haz. Aalen's lin. haz. regression (with censoring)
  - Parametric time series models
  - etc.

Challenges for the original FIC

- Local misspecification approach has been criticized
  - "Unrealistic theoretical framework"
  - "Not to yielding a valid environment for model selection"
  - Results cannot be trusted if the true model is far from the wide model
- The original FIC cannot compare:
  - non-nested parametric models
  - differently structured nonparametric and semiparametric type models

The new FIC construction strategy

- Motivated by the 'weaknesses' and critique of the original FIC
- Properties of the new FIC
  - Able to select between differently structured models (parametric vs. semi-/nonparametric models)
  - No need for parametric models to be nested or in any way related to each other
  - Derived without relying on a local misspecification framework
  - Carries with it an additional insurance mechanism against misspecification of (all) parametric models
  - May be viewed as multiple focused hypothesis test of the adequacy of the parametric models, with a theoretically chosen significance level
- Carried out for:
  - Parametric vs. nonparametric for i.i.d. data
  - Semiparametric Cox regression vs. fully parametric proportional regression
  - Parametric vs. nonparametric dependency modelling for time series

The new FIC construction strategy

- Motivated by the ‘weaknesses’ and critique of the original FIC
- Properties of the new FIC
  - Able to select between differently structured models (parametric vs. semi-/nonparametric models)
  - No need for parametric models to be nested or in any way related to each other
  - Derived without relying on a local misspecification framework
  - Carries with it an additional insurance mechanism against misspecification of (all) parametric models
  - May be viewed as multiple focused hypothesis test of the adequacy of the parametric models, with a theoretically chosen significance level
- Carried out for:
  - Parametric vs. nonparametric for i.i.d. data
  - Semiparametric Cox regression vs. fully parametric proportional regression
  - Parametric vs. nonparametric dependency modelling for time series

The new FIC construction strategy

- Motivated by the ‘weaknesses’ and critique of the original FIC
- Properties of the new FIC
  - Able to select between differently structured models (parametric vs. semi-/nonparametric models)
  - No need for parametric models to be nested or in any way related to each other
  - Derived without relying on a local misspecification framework
  - Carries with it an additional insurance mechanism against misspecification of (all) parametric models
  - May be viewed as multiple focused hypothesis test of the adequacy of the parametric models, with a theoretically chosen significance level
- Carried out for:
  - Parametric vs. nonparametric for i.i.d. data
  - Semiparametric Cox regression vs. fully parametric proportional regression
  - Parametric vs. nonparametric dependency modelling for time series

The new FIC for i.i.d. data

- Assume data $Y_1, \dots, Y_n$ stems from some fixed unknown distribution G
- Focus parameter defined through a functional mapping $\mu = T(\cdot)$ with $\mu_{\text{true}} = T(G)$
  - Typical focus parameters: Smooth functions of means and quantiles
- Estimate μ by plug-in estimation for each model S: $\widehat{\mu}_S = T(\widehat{G}_S)$
  - Nonparametric: $\widehat{\mu}_{\text{np}} = T(\widehat{G}_n)$ with $\widehat{G}_n$ the empirical distribution
  - Parametrics (generic): $\widehat{\mu}_{\text{pm}} = T(F_{\widehat{\theta}})$ with $\widehat{\theta}$ the MLE
- Performance measure:

$$\text{risk}_S = \text{mse}(\widehat{\mu}_S) = \text{E}\{(\widehat{\mu}_S - \mu_{\text{true}})^2\} = \text{bias}^2(\widehat{\mu}_S) + \text{var}(\widehat{\mu}_S)$$

- Approximate the risk by using asymptotic theory:

$$\widetilde{\text{risk}}_S = \widetilde{\text{mse}}(\widehat{\mu}_S) = \widetilde{\text{bias}}^2(\widehat{\mu}_S) + \widetilde{\text{var}}(\widehat{\mu}_S)$$

- The FIC score estimates the approximated risk

$$\text{FIC}_S = \widehat{\widetilde{\text{mse}}}(\widehat{\mu}_S) = \widehat{\widetilde{\text{bias}}^2}(\widehat{\mu}_S) + \widehat{\widetilde{\text{var}}}(\widehat{\mu}_S)$$

The new FIC for i.i.d. data

- Assume data $Y_1, \ldots, Y_n$ stems from some fixed unknown distribution G
- Focus parameter defined through a functional mapping $\mu = T(\cdot)$ with $\mu_{\rm true} = T(G)$
  - Typical focus parameters: Smooth functions of means and quantiles
- Estimate μ by plug-in estimation for each model S: $\widehat{\mu}_S = T(\widehat{G}_S)$
  - Nonparametric: $\widehat{\mu}_{\rm np} = T(\widehat{G}_n)$ with $\widehat{G}_n$ the empirical distribution
  - Parametrics (generic): $\widehat{\mu}_{\rm pm} = T(F_{\widehat{\theta}})$ with $\widehat{\theta}$ the MLE
- Performance measure:

$$\mathrm{risk}_S = \mathrm{mse}(\widehat{\mu}_S) = \mathrm{E}\left\{(\widehat{\mu}_S - \mu_{\rm true})^2\right\} = \mathrm{bias}^2(\widehat{\mu}_S) + \mathrm{var}(\widehat{\mu}_S)$$

- Approximate the risk by using asymptotic theory:

$$\widetilde{\mathrm{risk}}_S = \widetilde{\mathrm{mse}}(\widehat{\mu}_S) = \widetilde{\mathrm{bias}^2}(\widehat{\mu}_S) + \widetilde{\mathrm{var}}(\widehat{\mu}_S)$$

- The FIC score estimates the approximated risk

$$\mathrm{FIC}_S = \widehat{\widetilde{\mathrm{mse}}}(\widehat{\mu}_S) = \widehat{\widetilde{\mathrm{bias}^2}}(\widehat{\mu}_S) + \widehat{\widetilde{\mathrm{var}}}(\widehat{\mu}_S)$$

The new FIC for i.i.d. data

- Assume data $Y_1, \ldots, Y_n$ stems from some fixed unknown distribution G
- Focus parameter defined through a functional mapping $\mu = T(\cdot)$ with $\mu_{\mathrm{true}} = T(G)$
  - Typical focus parameters: Smooth functions of means and quantiles
- Estimate μ by plug-in estimation for each model S: $\widehat{\mu}_S = T(\widehat{G}_S)$
  - Nonparametric: $\widehat{\mu}_{\mathrm{np}} = T(\widehat{G}_n)$ with $\widehat{G}_n$ the empirical distribution
  - Parametrics (generic): $\widehat{\mu}_{\mathrm{pm}} = T(F_{\widehat{\theta}})$ with $\widehat{\theta}$ the MLE
- Performance measure:

$$\mathrm{risk}_S = \mathrm{mse}(\widehat{\mu}_S) = \mathrm{E}\left\{(\widehat{\mu}_S - \mu_{\mathrm{true}})^2\right\} = \mathrm{bias}^2(\widehat{\mu}_S) + \mathrm{var}(\widehat{\mu}_S)$$

- Approximate the risk by using asymptotic theory:

$$\widetilde{\mathrm{risk}}_S = \widetilde{\mathrm{mse}}(\widehat{\mu}_S) = \widetilde{\mathrm{bias}^2}(\widehat{\mu}_S) + \widetilde{\mathrm{var}}(\widehat{\mu}_S)$$

- The FIC score estimates the approximated risk

$$\mathrm{FIC}_S = \widehat{\widetilde{\mathrm{mse}}}(\widehat{\mu}_S) = \widehat{\widetilde{\mathrm{bias}^2}}(\widehat{\mu}_S) + \widehat{\widetilde{\mathrm{var}}}(\widehat{\mu}_S)$$

The new FIC for i.i.d. data

- Assume data $Y_1, \ldots, Y_n$ stems from some fixed unknown distribution G
- Focus parameter defined through a functional mapping $\mu = T(\cdot)$ with $\mu_{\mathrm{true}} = T(G)$
  - Typical focus parameters: Smooth functions of means and quantiles
- Estimate μ by plug-in estimation for each model S: $\widehat{\mu}_S = T(\widehat{G}_S)$
  - Nonparametric: $\widehat{\mu}_{\mathrm{np}} = T(\widehat{G}_n)$ with $\widehat{G}_n$ the empirical distribution
  - Parametrics (generic): $\widehat{\mu}_{\mathrm{pm}} = T(F_{\widehat{\theta}})$ with $\widehat{\theta}$ the MLE
- Performance measure:

$$\mathrm{risk}_S = \mathrm{mse}(\widehat{\mu}_S) = \mathrm{E}\left\{(\widehat{\mu}_S - \mu_{\mathrm{true}})^2\right\} = \mathrm{bias}^2(\widehat{\mu}_S) + \mathrm{var}(\widehat{\mu}_S)$$

- Approximate the risk by using asymptotic theory:

$$\widetilde{\mathrm{risk}}_S = \widetilde{\mathrm{mse}}(\widehat{\mu}_S) = \widetilde{\mathrm{bias}^2}(\widehat{\mu}_S) + \widetilde{\mathrm{var}}(\widehat{\mu}_S)$$

- The FIC score estimates the approximated risk

$$\mathrm{FIC}_S = \widehat{\widetilde{\mathrm{mse}}}(\widehat{\mu}_S) = \widehat{\widetilde{\mathrm{bias}^2}(\widehat{\mu}_S)} + \widehat{\widetilde{\mathrm{var}}(\widehat{\mu}_S)}$$

FIC construction: Derivation details (I)

- Let (generically) $\mu_{0,\mathrm{pm}} = T(F_{\theta_0})$ be the least false focus parameter value with θ_0 be the least false parameter value of θ

Under weak regularity conditions:

- Nonparametric: $\Psi_{\mathrm{np},n} = \sqrt{n}(\widehat{\mu}_{\mathrm{np}} - \mu_{\mathrm{true}}) \to_d \Psi_{\mathrm{np}} \sim \mathrm{N}(0, v_{\mathrm{np}})$
- Parametric $\Psi_{\mathrm{pm},n} = \sqrt{n}(\widehat{\mu}_{\mathrm{pm}} - \mu_{0,\mathrm{pm}}) \to_d \Psi_{\mathrm{pm}} \sim \mathrm{N}(0, v_{\mathrm{pm}})$

Large sample approximate risks

$$\text{Nonparametric: } \mathrm{mse}(\widehat{\mu}_{\mathrm{np}}) \approx \widetilde{\mathrm{mse}}(\widehat{\mu}_{\mathrm{np}}) = 0 + \frac{1}{n} v_{\mathrm{np}}$$

$$\text{Parametric: } \mathrm{mse}(\widehat{\mu}_{\mathrm{pm}}) \approx \widetilde{\mathrm{mse}}(\widehat{\mu}_{\mathrm{pm}}) = b^2 + \frac{1}{n} v_{\mathrm{pm}}$$

where $b = \mu_{0,\mathrm{pm}} - \mu_{\mathrm{true}}$

FIC construction: Derivation details (I)

- Let (generically) $\mu_{0,\mathrm{pm}} = T(F_{\theta_0})$ be the least false focus parameter value with θ_0 be the least false parameter value of θ

Under weak regularity conditions:

- Nonparametric: $\Psi_{\mathrm{np},n} = \sqrt{n}(\widehat{\mu}_{\mathrm{np}} - \mu_{\mathrm{true}}) \to_d \Psi_{\mathrm{np}} \sim \mathrm{N}(0, v_{\mathrm{np}})$
- Parametric $\Psi_{\mathrm{pm},n} = \sqrt{n}(\widehat{\mu}_{\mathrm{pm}} - \mu_{0,\mathrm{pm}}) \to_d \Psi_{\mathrm{pm}} \sim \mathrm{N}(0, v_{\mathrm{pm}})$

Large sample approximate risks

$$\text{Nonparametric: } \mathrm{mse}(\widehat{\mu}_{\mathrm{np}}) \approx \widetilde{\mathrm{mse}}(\widehat{\mu}_{\mathrm{np}}) = 0 + \frac{1}{n} v_{\mathrm{np}}$$

$$\text{Parametric: } \mathrm{mse}(\widehat{\mu}_{\mathrm{pm}}) \approx \widetilde{\mathrm{mse}}(\widehat{\mu}_{\mathrm{pm}}) = b^2 + \frac{1}{n} v_{\mathrm{pm}}$$

where $b = \mu_{0,\mathrm{pm}} - \mu_{\mathrm{true}}$

FIC construction: Derivation details (II)

- The convergence in distribution also holds jointly:

$$\sqrt{n}\begin{pmatrix}\widehat{\mu}_{\rm np}-\mu_{\rm true}\\ \widehat{\mu}_{\rm pm}-\mu_{0,\rm pm}\end{pmatrix}\to_d \mathrm{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}v_{\rm np} & v_{\rm c}\\ v_{\rm c} & v_{\rm pm}\end{pmatrix}\right)$$

- $\Rightarrow \sqrt{n}(\widehat{b}-b)\to_d \mathrm{N}(0,v_b)$, with $\widehat{b}=\widehat{\mu}_{\rm pm}-\widehat{\mu}_{\rm np}$ and $v_b=v_{\rm pm}+v_{\rm np}-2v_c$
- $\Rightarrow \mathsf{E}\{(\widehat{b})^2\}=(\mathsf{E}\{\widehat{b}\})^2+\mathsf{Var}(\widehat{b})\approx b^2+\frac{1}{n}v_b$

General estimates of risk components

- Squared parametric bias: b^2 estimated by $\max\{0,\widehat{b}^2-\frac{1}{n}\widehat{v}_b\}$
- Variances: $v_{\rm np}$ estimated by $\widehat{v}_{\rm np}$, and $v_{\rm pm}$ estimated by $\widehat{v}_{\rm pm}$

FIC scores

$$\text{Nonparametric: }\mathrm{FIC}(\widehat{\mu}_{\rm np})=\widehat{\mathrm{mse}}(\widehat{\mu}_{\rm np})=0+\frac{1}{n}\widehat{v}_{\rm np}$$

$$\text{Parametric: }\mathrm{FIC}(\widehat{\mu}_{\rm pm})=\widehat{\mathrm{mse}}(\widehat{\mu}_{\rm pm})=\max\{0,(\widehat{b})^2-\frac{1}{n}\widehat{v}_{\rm b}\}+\frac{1}{n}\widehat{v}_{\rm pm}$$

FIC construction: Derivation details (II)

- The convergence in distribution also holds jointly:

$$\sqrt{n}\begin{pmatrix}\widehat{\mu}_{\mathrm{np}}-\mu_{\mathrm{true}}\\ \widehat{\mu}_{\mathrm{pm}}-\mu_{0,\mathrm{pm}}\end{pmatrix}\to_d \mathrm{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}v_{\mathrm{np}} & v_{\mathrm{c}}\\ v_{\mathrm{c}} & v_{\mathrm{pm}}\end{pmatrix}\right)$$

- $\Rightarrow \sqrt{n}(\widehat{b}-b)\to_d \mathrm{N}(0,v_b)$, with $\widehat{b}=\widehat{\mu}_{\mathrm{pm}}-\widehat{\mu}_{\mathrm{np}}$ and $v_b=v_{\mathrm{pm}}+v_{\mathrm{np}}-2v_c$
- $\Rightarrow \mathsf{E}\{(\widehat{b})^2\}=(\mathsf{E}\{\widehat{b}\})^2+\mathsf{Var}(\widehat{b})\approx b^2+\frac{1}{n}v_b$

General estimates of risk components

- Squared parametric bias: b^2 estimated by $\max\{0,\widehat{b}^2-\frac{1}{n}\widehat{v}_b\}$
- Variances: v_{np} estimated by $\widehat{v}_{\mathrm{np}}$, and v_{pm} estimated by $\widehat{v}_{\mathrm{pm}}$

FIC scores

Nonparametric: $\mathrm{FIC}(\widehat{\mu}_{\mathrm{np}})=\widehat{\mathrm{mse}}(\widehat{\mu}_{\mathrm{np}})=0+\dfrac{1}{n}\widehat{v}_{\mathrm{np}}$

Parametric: $\mathrm{FIC}(\widehat{\mu}_{\mathrm{pm}})=\widehat{\mathrm{mse}}(\widehat{\mu}_{\mathrm{pm}})=\max\{0,(\widehat{b})^2-\dfrac{1}{n}\widehat{v}_{\mathrm{b}}\}+\dfrac{1}{n}\widehat{v}_{\mathrm{pm}}$

FIC construction: Derivation details (II)

- The convergence in distribution also holds jointly:

$$\sqrt{n}\begin{pmatrix}\widehat{\mu}_{\rm np}-\mu_{\rm true}\\ \widehat{\mu}_{\rm pm}-\mu_{0,{\rm pm}}\end{pmatrix}\to_d {\rm N}_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}v_{\rm np} & v_{\rm c}\\ v_{\rm c} & v_{\rm pm}\end{pmatrix}\right)$$

- $\Rightarrow \sqrt{n}(\widehat{b}-b)\to_d {\rm N}(0,v_b)$, with $\widehat{b}=\widehat{\mu}_{\rm pm}-\widehat{\mu}_{\rm np}$ and $v_b=v_{\rm pm}+v_{\rm np}-2v_c$
- $\Rightarrow \mathsf{E}\{(\widehat{b})^2\}=(\mathsf{E}\{\widehat{b}\})^2+\mathsf{Var}(\widehat{b})\approx b^2+\frac{1}{n}v_b$

General estimates of risk components

- Squared parametric bias: b^2 estimated by $\max\{0,\widehat{b}^2-\frac{1}{n}\widehat{v}_b\}$
- Variances: $v_{\rm np}$ estimated by $\widehat{v}_{\rm np}$, and $v_{\rm pm}$ estimated by $\widehat{v}_{\rm pm}$

FIC scores

Nonparametric: $\mathrm{FIC}(\widehat{\mu}_{\rm np})=\widehat{\rm mse}(\widehat{\mu}_{\rm np})=0+\dfrac{1}{n}\widehat{v}_{\rm np}$

Parametric: $\mathrm{FIC}(\widehat{\mu}_{\rm pm})=\widehat{\rm mse}(\widehat{\mu}_{\rm pm})=\max\{0,(\widehat{b})^2-\dfrac{1}{n}\widehat{v}_{\rm b}\}+\dfrac{1}{n}\widehat{v}_{\rm pm}$

Average/weighted FIC

- Generalisation of FIC, allowing selection with a (weighted) set of focus parameters (as for the original FIC)
- Performance measure:

$$\mathsf{risk} = \int \mathrm{mse}(\widehat{\mu}_S(t))\,\mathsf{d}W(t) = \int \mathsf{E}\left[\{\widehat{\mu}_S(t) - \mu_{\mathsf{true}}(t)\}^2\right]\mathsf{d}W(t)$$

- Average FIC (AFIC) score:

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

FIC/AFIC asymptotics

Consider one of the parametric model $S : F_\theta$ with least false model specification F_{θ_0} and least false focus parameter value $\mu_{0,\mathrm{pm}}$:

- If S is biased $(\mu_{\mathrm{true}} \neq \mu_{0,\mathrm{pm}})$
  - Pr (FIC/AFIC selects S) $\to 0$
- If S is fully correct $(G = F_{\theta_0})$
  - Pr (FIC selects S over nonparametric) $\to \chi^2_1(2) \approx 0.843$
  - No general results for AFIC (depends on μ and weight W)
  - AFIC for parametric with focus on all $\mu(y) = G(y)$ and $W = F_{\theta_0}$
    - Re-invents the Cramér-von Mises goodness-of-fit test, with a threshold value found by the theory itself

$\Rightarrow$ If all parametric candidate models are biased:
Pr (FIC/AFIC selects nonparametric) $\to 1$

- Insurance mechanism against parametric misspecification

FIC/AFIC asymptotics

Consider one of the parametric model $S : F_\theta$ with least false model specification F_{θ_0} and least false focus parameter value $\mu_{0,\mathrm{pm}}$:

- If S is biased $(\mu_{\mathrm{true}} \neq \mu_{0,\mathrm{pm}})$
  - $\Pr\left(\text{FIC/AFIC selects } S\right) \to 0$
- If S is fully correct $(G = F_{\theta_0})$
  - $\Pr\left(\text{FIC selects } S \text{ over nonparametric}\right) \to \chi^2_1(2) \approx 0.843$
  - No general results for AFIC (depends on μ and weight W)
  - AFIC for parametric with focus on all $\mu(y) = G(y)$ and $W = F_{\theta_0}$
    - Re-invents the Cramér-von Mises goodness-of-fit test, with a threshold value found by the theory itself

$\Rightarrow$ If all parametric candidate models are biased: $\Pr\left(\text{FIC/AFIC selects nonparametric}\right) \to 1$

- Insurance mechanism against parametric misspecification

FIC/AFIC asymptotics

Consider one of the parametric model $S : F_\theta$ with least false model specification F_{θ_0} and least false focus parameter value $\mu_{0,\mathrm{pm}}$:

- If S is biased $(\mu_{\mathrm{true}} \neq \mu_{0,\mathrm{pm}})$
    - $\mathrm{Pr}\,(\text{FIC/AFIC selects } S) \to 0$
- If S is fully correct $(G = F_{\theta_0})$
    - $\mathrm{Pr}\,(\text{FIC selects } S \text{ over nonparametric}) \to \chi^2_1(2) \approx 0.843$
    - No general results for AFIC (depends on μ and weight W)
    - AFIC for parametric with focus on all $\mu(y) = G(y)$ and $W = F_{\theta_0}$
        - Re-invents the Cramér-von Mises goodness-of-fit test, with a threshold value found by the theory itself

$\Rightarrow$ If all parametric candidate models are biased:
$\mathrm{Pr}\,(\text{FIC/AFIC selects nonparametric}) \to 1$

- Insurance mechanism against parametric misspecification

Other data types (I)

Censoring

- Semiparametric Cox regression vs. fully parametric proportional hazard regression models
  - $G = (A(\cdot), \beta)$ corresponding to the hazard rate function: $\alpha(s)\exp(x^{\mathrm{t}}\beta)$, with α unspecified baseline hazard and A its cumulative
  - Semiparametric Cox regression: $\widehat{\beta}_{\mathsf{cox}}$ via Cox's partial likelihood, $\widehat{A}_{\mathsf{cox}}(\cdot) =$ Breslow estimator
  - Parametric estimation: Joint ML estimation of θ, β, with a parametric hazard rate function α_θ (exponential, Weibull, Gompertz, ...)
  - Typical focus parameters: Survival probabilities, quantiles and cumulative hazards, conditional on covariate values

- Without covariates: Nelson–Aalen or Kaplan–Meier estimators vs. parametric survival models

Other data types (II)

Dependency modelling in stationary Gaussian time series models

- $G =$ Unknown spectral measure/distribution
- Nonparametric estimation: Integrated periodogram $\widehat{G}_n(\omega) = \int_{-\pi}^{\omega} I_n(u)\,\mathrm{d}u$ for I_n the periodogram
- Parametric estimation: Typically ARMA-models, but also more general parametric models for the spectral distribution
- Typical focus parameters: Differences in spectral distribution, covariance lags and correlation lags

Slightly different strategies

- Density estimation for i.i.d. data
- Standard regression

Other data types (II)

Dependency modelling in stationary Gaussian time series models

- $G =$ Unknown spectral measure/distribution
- Nonparametric estimation: Integrated periodogram $\widehat{G}_n(\omega) = \int_{-\pi}^{\omega} I_n(u)\,\mathrm{d}u$ for I_n the periodogram
- Parametric estimation: Typically ARMA-models, but also more general parametric models for the spectral distribution
- Typical focus parameters: Differences in spectral distribution, covariance lags and correlation lags

Slightly different strategies

- Density estimation for i.i.d. data
- Standard regression

Original vs. new FIC scores

- Assume for comparison that all parametric models are nested between a wide (θ, γ both free) and narrow (θ free, γ fixed) model
- Original FIC scores:

$$\mathrm{FIC}_{\text{wide,orig}} = \widehat{v}_{\text{wide}}$$

$$\mathrm{FIC}_{S,\text{orig}} = \max\{0, \widehat{Z}_S^2 - (\widehat{v}_{\text{wide}} - \widehat{v}_S)\} + \widehat{v}_S$$

- $\widehat{Z}_S$ is the natural estimator of the bias of $\Lambda_S \overset{d}{=} \lim_{n\to\infty} \sqrt{n}(\widehat{\mu}_S - \mu_{\text{true},n})$ ($\widehat{Z}_S = \widehat{\omega}^{\mathrm{t}}(I_q - \widehat{G}_S)D_n$ in C&H (2008) notation)
- New FIC scores (scaled by n to ease comparison):

$$\mathrm{FIC}_{\text{np,new}} = \widehat{v}_{\text{np}}$$

$$\mathrm{FIC}_{\text{wide,new}} = \max\{0, (\sqrt{n}\widehat{b}_{\text{wide}})^2 - (\widehat{v}_{\text{np}} + \widehat{v}_{\text{wide,new}} - 2\widehat{v}_{c,\text{wide}})\} + \widehat{v}_{\text{wide,new}}$$

$$\mathrm{FIC}_{S,\text{new}} = \max\{0, (\sqrt{n}\widehat{b}_S)^2 - (\widehat{v}_{\text{np}} + \widehat{v}_{S,\text{new}} - 2\widehat{v}_{c,S})\} + \widehat{v}_{S,\text{new}}$$

Original vs. new FIC scores

- Assume for comparison that all parametric models are nested between a wide (θ, γ both free) and narrow (θ free, γ fixed) model
- Original FIC scores:

$$\mathrm{FIC}_{\mathrm{wide,orig}} = \widehat{v}_{\mathrm{wide}}$$
$$\mathrm{FIC}_{S,\mathrm{orig}} = \max\{0, \widehat{Z}_S^2 - (\widehat{v}_{\mathrm{wide}} - \widehat{v}_S)\} + \widehat{v}_S$$

- $\widehat{Z}_S$ is the natural estimator of the bias of $\Lambda_S \overset{d}{=} \lim_{n\to\infty} \sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n})$ ($\widehat{Z}_S = \widehat{\omega}^{\mathrm{t}}(I_q - \widehat{G}_S)D_n$ in C&H (2008) notation)
- New FIC scores (scaled by n to ease comparison):

$$\mathrm{FIC}_{\mathrm{np,new}} = \widehat{v}_{\mathrm{np}}$$
$$\mathrm{FIC}_{\mathrm{wide,new}} = \max\{0, (\sqrt{n}\widehat{b}_{\mathrm{wide}})^2 - (\widehat{v}_{\mathrm{np}} + \widehat{v}_{\mathrm{wide,new}} - 2\widehat{v}_{c,\mathrm{wide}})\} + \widehat{v}_{\mathrm{wide,new}}$$
$$\mathrm{FIC}_{S,\mathrm{new}} = \max\{0, (\sqrt{n}\widehat{b}_S)^2 - (\widehat{v}_{\mathrm{np}} + \widehat{v}_{S,\mathrm{new}} - 2\widehat{v}_{c,S})\} + \widehat{v}_{S,\mathrm{new}}$$

Original vs. new FIC scores

- Assume for comparison that all parametric models are nested between a wide (θ, γ both free) and narrow (θ free, γ fixed) model
- Original FIC scores:

$$\mathrm{FIC}_{\mathrm{wide,orig}} = \widehat{v}_{\mathrm{wide}}$$
$$\mathrm{FIC}_{S,\mathrm{orig}} = \max\{0, \widehat{Z}_S^2 - (\widehat{v}_{\mathrm{wide}} - \widehat{v}_S)\} + \widehat{v}_S$$

- $\widehat{Z}_S$ is the natural estimator of the bias of $\Lambda_S \overset{d}{=} \lim_{n\to\infty} \sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true},n})$ ($\widehat{Z}_S = \widehat{\omega}^{\mathrm{t}}(I_q - \widehat{G}_S)D_n$ in C&H (2008) notation)
- New FIC scores (scaled by n to ease comparison):

$$\mathrm{FIC}_{\mathrm{np,new}} = \widehat{v}_{\mathrm{np}}$$
$$\mathrm{FIC}_{\mathrm{wide,new}} = \max\{0, (\sqrt{n}\widehat{b}_{\mathrm{wide}})^2 - (\widehat{v}_{\mathrm{np}} + \widehat{v}_{\mathrm{wide,new}} - 2\widehat{v}_{c,\mathrm{wide}})\} + \widehat{v}_{\mathrm{wide,new}}$$
$$\mathrm{FIC}_{S,\mathrm{new}} = \max\{0, (\sqrt{n}\widehat{b}_S)^2 - (\widehat{v}_{\mathrm{np}} + \widehat{v}_{S,\mathrm{new}} - 2\widehat{v}_{c,S})\} + \widehat{v}_{S,\mathrm{new}}$$

Asymptotic comp. under local misspecification

Limit of original FIC scores

$$\mathrm{FIC}_{\mathrm{wide,orig}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{wide,orig}} = v_{\mathrm{wide}},$$

$$\mathrm{FIC}_{S,\mathrm{orig}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{S,\mathrm{orig}} = \max\{0, Z_S^2 - (v_{\mathrm{wide}} - v_S)\} + v_S$$

- $Z_S = \lim_{n\to\infty} \widehat{Z}_S \quad (= \omega^{\mathrm{t}}(I_q - G_S)D$ in C&H (2008) notation)

Limit of new FIC scores

$$\mathrm{FIC}_{\mathrm{np,new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{np,new}} = v_{\mathrm{np}},$$

$$\mathrm{FIC}_{\mathrm{wide,new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{wide,new}} = \max\{0, Z_0^2 - (v_{\mathrm{np}} - v_{\mathrm{wide}})\} + v_{\mathrm{wide}},$$

$$\mathrm{FIC}_{S,\mathrm{new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{S,\mathrm{new}} = \max\{0, (Z_S - Z_0)^2 - (v_{\mathrm{np}} - v_S)\} + v_S$$

- $Z_0 \overset{d}{=} \lim_{n\to\infty} \sqrt{n}(\widehat{\mu}_{\mathrm{np}} - \widehat{\mu}_{\mathrm{wide}})$, is zero-mean normal and stochastically independent of all Z_S
- New FIC has an additional uncertainty level (Z_0), uses nonparametric (model robust) estimates instead of wide model variances
- Original and new FIC asymptotically equivalent when $v_{\mathrm{np}} = v_{\mathrm{wide}}$

Asymptotic comp. under local misspecification

Limit of original FIC scores

$$\mathrm{FIC}_{\mathrm{wide,orig}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{wide,orig}} = v_{\mathrm{wide}},$$

$$\mathrm{FIC}_{S,\mathrm{orig}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{S,\mathrm{orig}} = \max\{0, Z_S^2 - (v_{\mathrm{wide}} - v_S)\} + v_S$$

- $Z_S = \lim_{n\to\infty} \widehat{Z}_S \quad (= \omega^{\mathrm{t}}(I_q - G_S)D$ in C&H (2008) notation)

Limit of new FIC scores

$$\mathrm{FIC}_{\mathrm{np,new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{np,new}} = v_{\mathrm{np}},$$

$$\mathrm{FIC}_{\mathrm{wide,new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{wide,new}} = \max\{0, Z_0^2 - (v_{\mathrm{np}} - v_{\mathrm{wide}})\} + v_{\mathrm{wide}},$$

$$\mathrm{FIC}_{S,\mathrm{new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{S,\mathrm{new}} = \max\{0, (Z_S - Z_0)^2 - (v_{\mathrm{np}} - v_S)\} + v_S$$

- $Z_0 \overset{d}{=} \lim_{n\to\infty} \sqrt{n}(\widehat{\mu}_{\mathrm{np}} - \widehat{\mu}_{\mathrm{wide}})$, is zero-mean normal and stochastically independent of all Z_S
- New FIC has an additional uncertainty level (Z_0), uses nonparametric (model robust) estimates instead of wide model variances
- Original and new FIC asymptotically equivalent when $v_{\mathrm{np}} = v_{\mathrm{wide}}$

Asymptotic comp. under local misspecification

Limit of original FIC scores

$$\mathrm{FIC}_{\mathrm{wide,orig}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{wide,orig}} = v_{\mathrm{wide}},$$

$$\mathrm{FIC}_{S,\mathrm{orig}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{S,\mathrm{orig}} = \max\{0, Z_S^2 - (v_{\mathrm{wide}} - v_S)\} + v_S$$

- $Z_S = \lim_{n\to\infty} \widehat{Z}_S \quad (= \omega^{\mathrm{t}}(I_q - G_S)D$ in C&H (2008) notation)

Limit of new FIC scores

$$\mathrm{FIC}_{\mathrm{np,new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{np,new}} = v_{\mathrm{np}},$$

$$\mathrm{FIC}_{\mathrm{wide,new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{\mathrm{wide,new}} = \max\{0, Z_0^2 - (v_{\mathrm{np}} - v_{\mathrm{wide}})\} + v_{\mathrm{wide}},$$

$$\mathrm{FIC}_{S,\mathrm{new}} \xrightarrow{d} \mathrm{FIC}^{\mathrm{lim}}_{S,\mathrm{new}} = \max\{0, (Z_S - Z_0)^2 - (v_{\mathrm{np}} - v_S)\} + v_S$$

- $Z_0 \stackrel{d}{=} \lim_{n\to\infty} \sqrt{n}(\widehat{\mu}_{\mathsf{np}} - \widehat{\mu}_{\mathsf{wide}})$, is zero-mean normal and stochastically independent of all Z_S
- New FIC has an additional uncertainty level (Z_0), uses nonparametric (model robust) estimates instead of wide model variances
- Original and new FIC asymptotically equivalent when $v_{\mathrm{np}} = v_{\mathsf{wide}}$

Summary and concluding remarks

- New FIC construction strategy attempting to avoid 'weaknesses' of the original approach
    - Do not rely on local asymptotics
    - Able to handle non-nested parametric models
    - Allow comparison of parametrics and semi-/nonparametrics
- May be viewed as model robust version of the original FIC
- Carries with it an insurance mechanism against parametric misspecification
- Developed for a wide range of data types where few alternative model selection procedures are available
- Simulation experiments indicate performance better than AIC/BIC for moderate to large n
- Model averaging scheme also developed
- Possibly more difficult to generalize than the original FIC