New focused approaches to topics within model selection and approximate Bayesian inversion

Thesis presented for the degree of Philosophiae Doctor (PhD)

Martin Jullum

University of Oslo

martinju@math.uio.no

April 1, 2016

Thesis papers

Paper I

JULLUM, M. & HJORT, N. L. (2016). Parametric or nonparametric: The FIC approach. *Minor revision submitted for publication in Statistica Sinica*

Paper II

JULLUM, M. & HJORT, N. L. (2015). What price semiparametric Cox regression? *Submitted for publication in Scandinavian Journal of Statistics*

Paper III

HERMANSEN, G. H., HJORT, N. L. & JULLUM, M. (2015). Parametric or nonparametric: The FIC approach for stationary time series. *Technical report, Department of Mathematics, University of Oslo*

Paper IV

JULLUM, M. & KOLBJØRNSEN, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics* **81**(3), R1–R13.

Focused inference for the layman

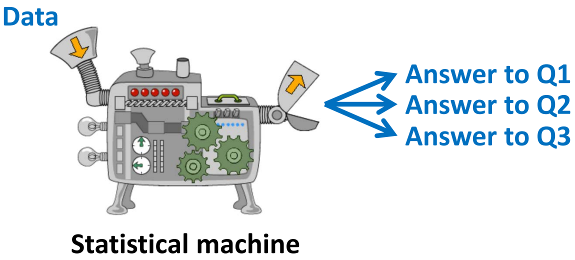
Data: Quantified information gathered from recordings, measurements or surveys

Statistics: Answer scientific questions under uncertainty based on mathematical modelling and analysis of data

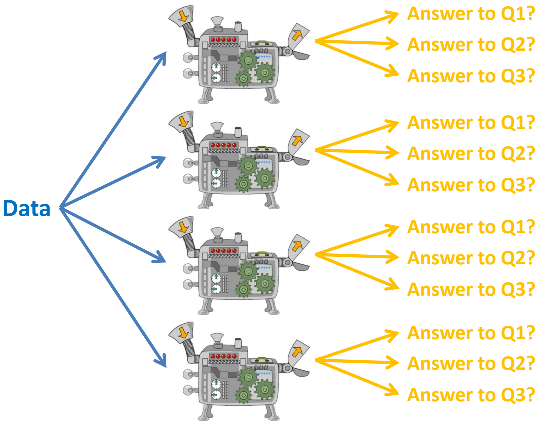
Focused inference for the layman

Data: Quantified information gathered from recordings, measurements or surveys

Statistics: Answer scientific questions under uncertainty based on mathematical modelling and analysis of data

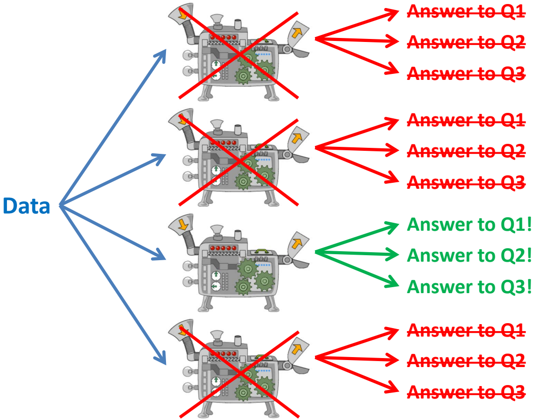


Focused inference for the layman



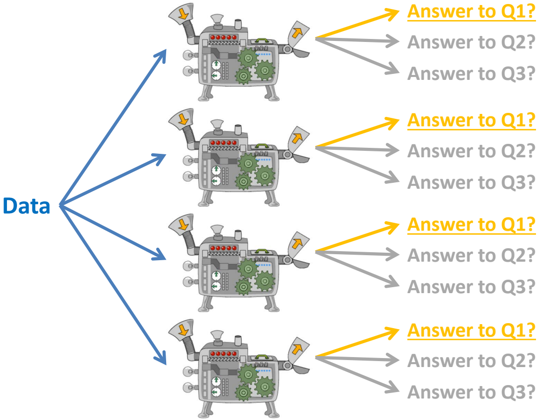
Focused inference for the layman

Traditional (unfocused) approach



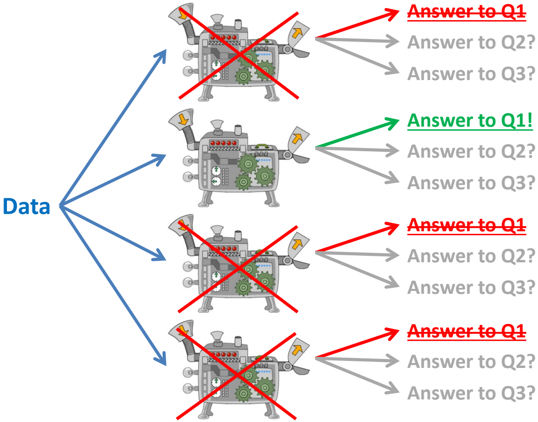
Focused inference for the layman

Focused approach (question 1)



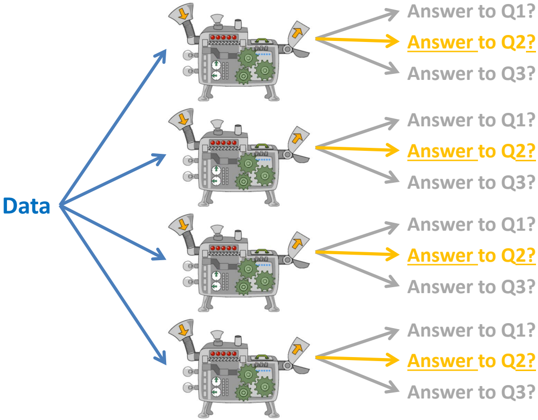
Focused inference for the layman

Focused approach (question 1)



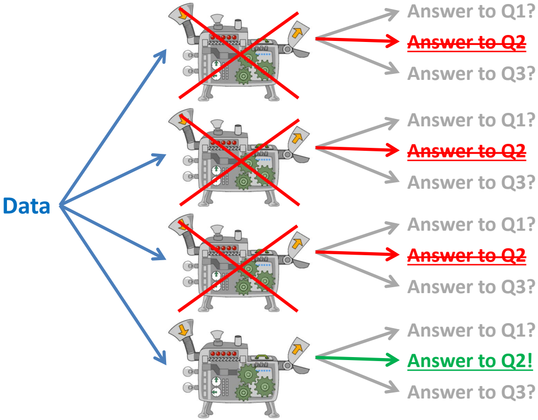
Focused inference for the layman

Focused approach (question 2)



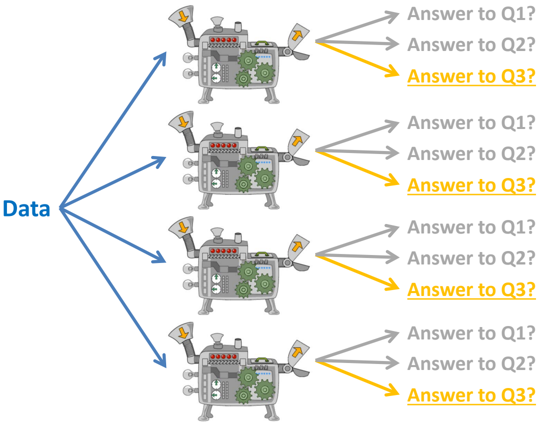
Focused inference for the layman

Focused approach (question 2)



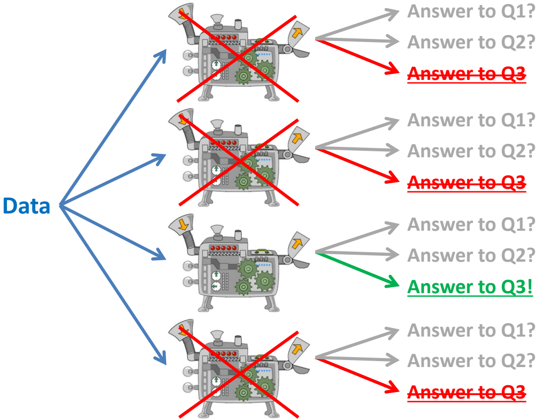
Focused inference for the layman

Focused approach (question 3)



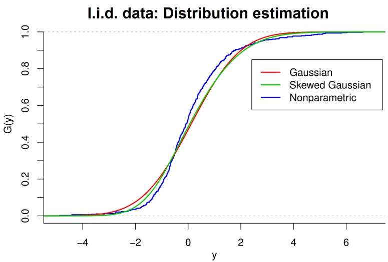
Focused inference for the layman

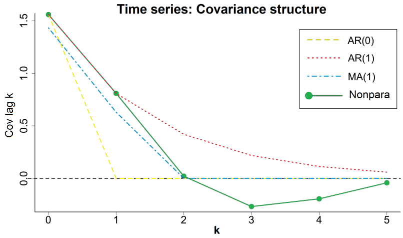
Focused approach (question 3)

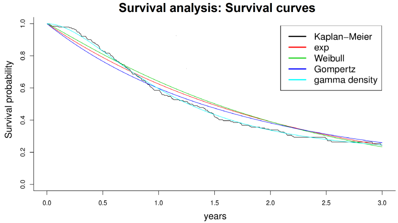


Model selection

- Unknown data generating mechanism G_0 for data $Y_1, \ldots, Y_n$
- Parametric approaches
- Nonparametric approach
- Several appropriate models – which one should we trust?







Motivation and research objective

- Traditional model selection approaches (AIC, BIC, DIC,...) cannot handle 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

Main research objective Papers I-III

Construct focused/interest driven model selection criteria for selection among a set of parametric and nonparametric type models

Motivation and research objective

- Traditional model selection approaches (AIC, BIC, DIC,...) cannot handle 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

Main research objective Papers I-III

Construct focused/interest driven model selection criteria for selection among a set of parametric and nonparametric type models

Motivation and research objective

- Traditional model selection approaches (AIC, BIC, DIC,...) cannot handle 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

Main research objective Papers I-III

Construct focused/interest driven model selection criteria for selection among a set of parametric and nonparametric type models

Focus parameter

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

Examples

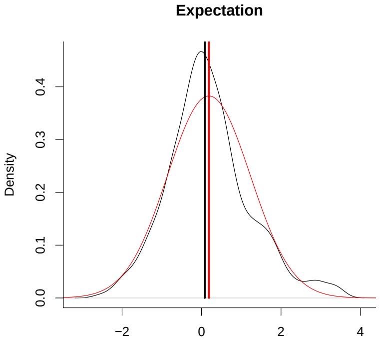
- Expectation: $\mu = T(G) = \mathrm{E}_G\left(Y_i\right)$
- $\Pr\left(Y_i > 2\right)$: $\mu = T(G) = 1 - G(2)$
- Interquartile range: $\mu = T(G) = G^{-1}(3/4) - G^{-1}(1/4)$

Focus parameter

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

Examples

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

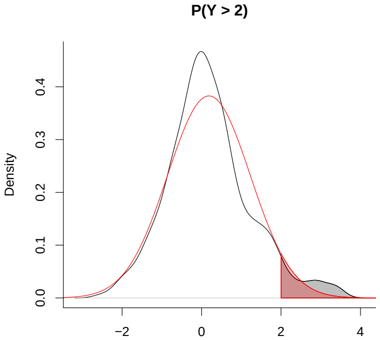


Focus parameter

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

Examples

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

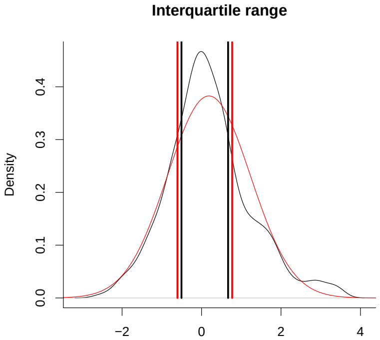


Focus parameter

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

Examples

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



Criterion idea

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

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

Basic idea: Focused information criterion (FIC)

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

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

- Choose the model and estimator with the smallest FIC score

Criterion idea

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

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

Basic idea: Focused information criterion (FIC)

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

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

- Choose the model and estimator with the smallest FIC score

Average/weighted FIC

- Generalisation of FIC for a (weighted) set of focus parameters
- Performance measure:
  risk $= \int \mathrm{mse}(\widehat{\mu}_M(t))\,\mathsf{d}W(t) = \int \mathsf{E}\left[\{\widehat{\mu}_M(t) - \mu_{\mathsf{true}}(t)\}^2\right]\mathsf{d}W(t)$,
- $\mathsf{AFIC}(M) = \int \widehat{\mathrm{mse}}(\widehat{\mu}_M(t))\,\mathsf{d}W(t) = \int \mathrm{FIC}(\widehat{\mu}_M(t))\,\mathsf{d}W(t)$

FIC/AFIC asymptotics (pm vs. np)

- Parametric model biased: Pr (FIC/AFIC selects pm) $\to 0$
- Parametric model correct: Pr (FIC selects pm) $\to \chi_1^2(2) \approx 0.843$

Original FIC

- Idea based on the original FIC by Claeskens & Hjort (2003)
- Our approach does not require a local misspecification framework and works for nonparametrics and with non-nested parametric models

Average/weighted FIC

- Generalisation of FIC for a (weighted) set of focus parameters
- Performance measure:
  risk $= \int \mathrm{mse}(\widehat{\mu}_M(t))\,\mathsf{d}W(t) = \int \mathsf{E}\left[\{\widehat{\mu}_M(t) - \mu_{\mathsf{true}}(t)\}^2\right]\mathsf{d}W(t)$,
- $\mathsf{AFIC}(M) = \int \widehat{\mathrm{mse}}(\widehat{\mu}_M(t))\,\mathsf{d}W(t) = \int \mathrm{FIC}(\widehat{\mu}_M(t))\,\mathsf{d}W(t)$

FIC/AFIC asymptotics (pm vs. np)

- Parametric model biased: Pr (FIC/AFIC selects pm) $\to 0$
- Parametric model correct: Pr (FIC selects pm) $\to \chi_1^2(2) \approx 0.843$

Original FIC

- Idea based on the original FIC by Claeskens & Hjort (2003)
- Our approach does not require a local misspecification framework and works for nonparametrics and with non-nested parametric models

Average/weighted FIC

- Generalisation of FIC for a (weighted) set of focus parameters
- Performance measure:
  risk $= \int \mathrm{mse}(\widehat{\mu}_M(t))\,\mathsf{d}W(t) = \int \mathsf{E}\left[\{\widehat{\mu}_M(t) - \mu_{\mathsf{true}}(t)\}^2\right]\mathsf{d}W(t)$,
- AFIC$(M) = \int \widehat{\mathrm{mse}}(\widehat{\mu}_M(t))\,\mathsf{d}W(t) = \int \mathrm{FIC}(\widehat{\mu}_M(t))\,\mathsf{d}W(t)$

FIC/AFIC asymptotics (pm vs. np)

- Parametric model biased: Pr (FIC/AFIC selects pm) $\to 0$
- Parametric model correct: Pr (FIC selects pm) $\to \chi^2_1(2) \approx 0.843$

Original FIC

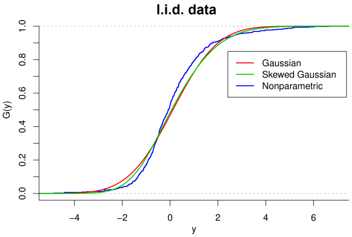
- Idea based on the original FIC by Claeskens & Hjort (2003)
- Our approach does not require a local misspecification framework and works for nonparametrics and with non-nested parametric models

Paper I

JULLUM, M. & HJORT, N. L. (2016). Parametric or nonparametric: The FIC approach. *Minor revision submitted for publication in Statistica Sinica*

- Main contribution: Develop and study the FIC construction routine for i.i.d. data
- G: The cumulative distribution function
- Nonparametric estimation: Empirical distribution function $\widehat{G}_n$
- Parametric estimation: Ordinary maximum likelihood estimation for parametric families F_θ

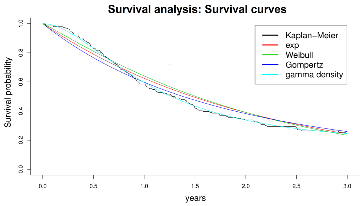
- Typical focus parameters: Smooth functions of means and quantiles
- Also discuss corresponding FIC schemes for density estimation and regression



Paper II

JULLUM, M. & HJORT, N. L. (2015). What price semiparametric Cox regression? *Submitted for publication in Scandinavian Journal of Statistics*

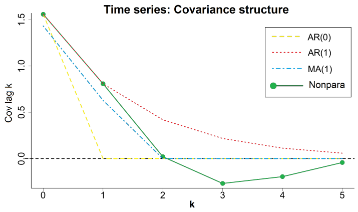
- Main contribution: Lifting the FIC framework to censored survival time data with covariates
- $G = \{A(\cdot), \beta\}$ corresponding to the hazard rate function: $\alpha(\cdot)\exp(x^{\mathrm{t}}\beta)$
- 'Nonparametric' estimation: Semiparametric Cox regression
- Parametric estimation: Joint ML estimation of θ, β, with a parametric hazard rate function $\alpha_\theta(\cdot)$
- Typical focus parameters: Survival probabilities, quantiles and cumulative hazards, conditional on covariate values
- Also investigate the asymptotic relative efficiency (ARE) for various focus parameters



Paper III

HERMANSEN, G. H., HJORT, N. L. & JULLUM, M. (2015). Parametric or nonparametric: The FIC approach for stationary time series. *Technical report, Department of Mathematics, University of Oslo*

- Main contribution: Lifts the FIC framework to stationary Gaussian time series
- G: Spectral measure/distribution
- Nonparametric estimation: Periodogram $\widehat{G}_n$
- Typical parametric alternatives: Autoregressive and moving average models, estimated by ML or using the Whittle approximation
- Typical focus parameters: Differences in spectral distribution, covariance lags and correlation lags



Forward and inverse problems

- Consider

$$y = H(x) + \varepsilon$$

  - y: observable data
  - x: latent cause/source
  - H: (causal) mechanism operator
  - ε: noise term

- Forward problem: 'Finding' y based on x
- Inverse problem: 'Finding' x based on y

Bayesian solution to the inverse problem

- Apply Bayes' formula $p(x|y) \propto p(y|x)p(x)$
- Consult posterior distribution $p(x|y)$

Forward and inverse problems

- Consider

$$y = H(x) + \varepsilon$$

  - y: observable data
  - x: latent cause/source
  - H: (causal) mechanism operator
  - ε: noise term

- Forward problem: 'Finding' y based on x
- Inverse problem: 'Finding' x based on y

Bayesian solution to the inverse problem

- Apply Bayes' formula $p(x|y) \propto p(y|x)p(x)$
- Consult posterior distribution $p(x|y)$

Forward and inverse problems

- Consider

$$y = H(x) + \varepsilon$$

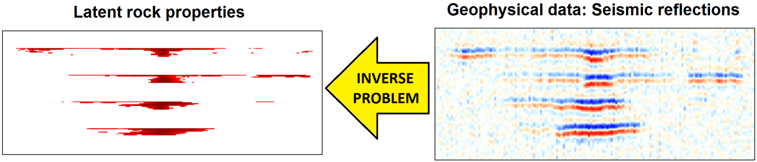
  - y: observable data
  - x: latent cause/source
  - H: (causal) mechanism operator
  - ε: noise term

- Forward problem: 'Finding' y based on x
- Inverse problem: 'Finding' x based on y

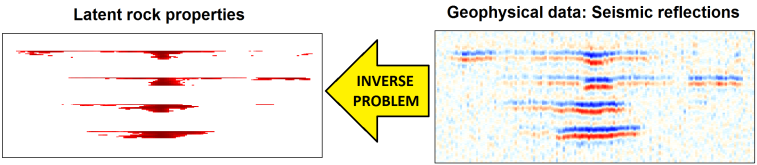
Bayesian solution to the inverse problem

- Apply Bayes' formula $p(x|y) \propto p(y|x)p(x)$
- Consult posterior distribution $p(x|y)$

Inverse problem within the geosciences



Inverse problem within the geosciences

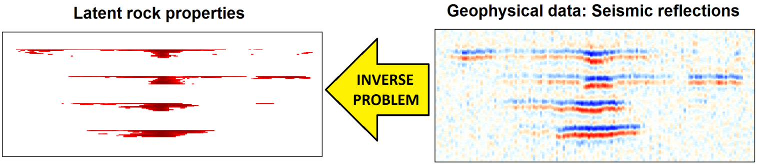


Forward model

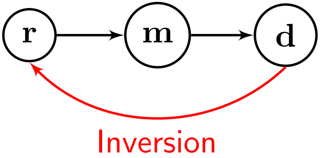
$$\mathbf{r} \longrightarrow \mathbf{m} \longrightarrow \mathbf{d}$$

- $\mathbf{r}$: Latent rock properties
- $\mathbf{m}$: Latent geophysical properties
- $\mathbf{d}$: Geophysical data

Inverse problem within the geosciences

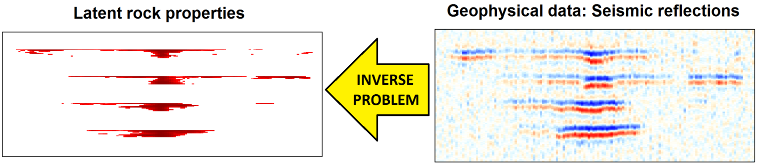


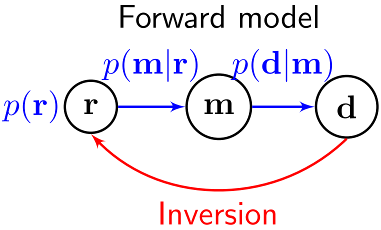
Forward model



- $\mathbf{r}$: Latent rock properties
- $\mathbf{m}$: Latent geophysical properties
- $\mathbf{d}$: Geophysical data

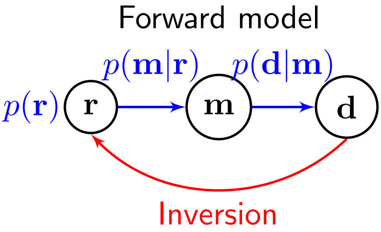
Inverse problem within the geosciences





- $\mathbf{r}$: Latent rock properties
- $\mathbf{m}$: Latent geophysical properties
- $\mathbf{d}$: Geophysical data

Main problem

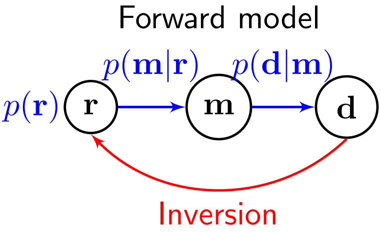


- $\mathbf{r}$: Latent rock properties
- $\mathbf{m}$: Latent geophysical properties
- $\mathbf{d}$: Geophysical data

Posterior distribution: $p(\mathbf{r}|\mathbf{d}) \propto \int p(\mathbf{d}|\mathbf{m})p(\mathbf{m}|\mathbf{r})p(\mathbf{r})\,d\mathbf{m}$

- High dimensional problem
- Enormous amount of highly correlated data $\mathbf{d}$
- Complex dependency structures
- Analytical expression for posterior seldom available
- MCMC can be very time consuming

Main problem



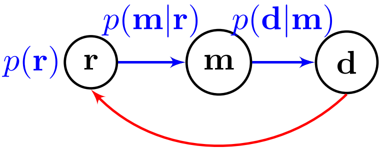
- $\mathbf{r}$: Latent rock properties
- $\mathbf{m}$: Latent geophysical properties
- $\mathbf{d}$: Geophysical data

Posterior distribution: $p(\mathbf{r}|\mathbf{d}) \propto \int p(\mathbf{d}|\mathbf{m})p(\mathbf{m}|\mathbf{r})p(\mathbf{r})\, d\mathbf{m}$

- High dimensional problem
- Enormous amount of highly correlated data $\mathbf{d}$
- Complex dependency structures
- Analytical expression for posterior seldom available
- MCMC can be very time consuming

Main problem

Forward model



Inversion

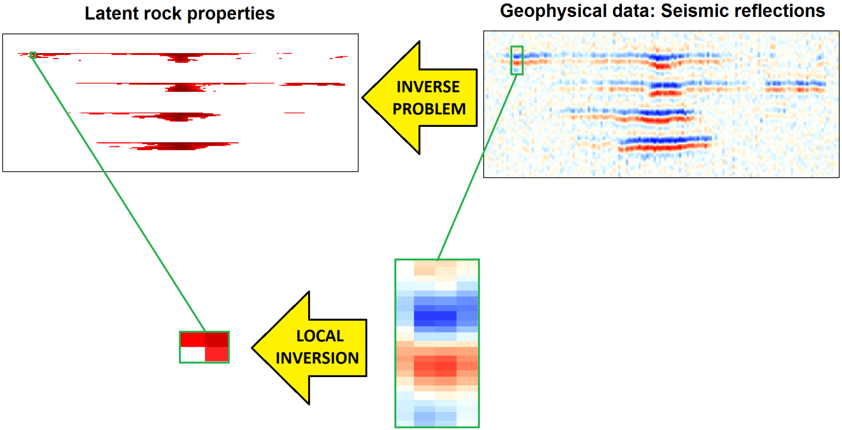
- $\mathbf{r}$: Latent rock properties
- $\mathbf{m}$: Latent geophysical properties
- $\mathbf{d}$: Geophysical data

Posterior distribution: $p(\mathbf{r}|\mathbf{d}) \propto \int p(\mathbf{d}|\mathbf{m})p(\mathbf{m}|\mathbf{r})p(\mathbf{r})\,\mathrm{d}\mathbf{m}$

- High dimensional problem
- Enormous amount of highly correlated data $\mathbf{d}$
- Complex dependency structures
- Analytical expression for posterior seldom available
- MCMC can be very time consuming

Paper IV

JULLUM, M. & KOLBJØRNSEN, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics* **81**(3), R1–R13.



Paper IV

JULLUM, M. & KOLBJØRNSEN, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics* **81**(3), R1–R13.

- Divide the global inversion problem into several local inversions
  - Approximate marginal posterior $p(\mathbf{r}_i|\mathbf{d})$ for each cell i in the gridded region, rather than the global $p(\mathbf{r}|\mathbf{d})$ for the full region
- Dimension reduction by only using variables spatially close to the cell in focus

- 3 neighborhoods of cells B, C, D, with local variables $\mathbf{r}_B, \mathbf{m}_C, \mathbf{d}_D$
- Approximate marginal posterior:

$$p(\mathbf{r}_i|\mathbf{d}_D) \approx p^*(\mathbf{r}_i|\mathbf{d}_D) \propto \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B_{-i}},$$

  with Gaussian approximation

  $p^*(\mathbf{d}_D|\mathbf{r}_B) = \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$
- Weighted Monte Carlo routine for sample based evaluation of posterior

Paper IV

JULLUM, M. & KOLBJØRNSEN, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics* **81**(3), R1–R13.

- Divide the global inversion problem into several local inversions
  - Approximate marginal posterior $p(\mathbf{r}_i|\mathbf{d})$ for each cell i in the gridded region, rather than the global $p(\mathbf{r}|\mathbf{d})$ for the full region
- Dimension reduction by only using variables spatially close to the cell in focus

Inversion for cell i: $p(\mathbf{r}_i|\mathbf{d})$

- 3 neighborhoods of cells B, C, D, with local variables $\mathbf{r}_B, \mathbf{m}_C, \mathbf{d}_D$
- Approximate marginal posterior:

$$p(\mathbf{r}_i|\mathbf{d}_D) \approx p^*(\mathbf{r}_i|\mathbf{d}_D) \propto \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B_{-i}},$$

  with Gaussian approximation
  $p^*(\mathbf{d}_D|\mathbf{r}_B) = \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$
- Weighted Monte Carlo routine for sample based evaluation of posterior

Paper IV

JULLUM, M. & KOLBJØRNSEN, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics* **81**(3), R1–R13.

- Divide the global inversion problem into several local inversions
  - Approximate marginal posterior $p(\mathbf{r}_i|\mathbf{d})$ for each cell i in the gridded region, rather than the global $p(\mathbf{r}|\mathbf{d})$ for the full region
- Dimension reduction by only using variables spatially close to the cell in focus

Inversion for cell i: $p(\mathbf{r}_i|\mathbf{d})$

- 3 neighborhoods of cells B, C, D, with local variables $\mathbf{r}_B, \mathbf{m}_C, \mathbf{d}_D$
- Approximate marginal posterior:

$$p(\mathbf{r}_i|\mathbf{d}_D) \approx p^*(\mathbf{r}_i|\mathbf{d}_D) \propto \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B_{-i}},$$

  with Gaussian approximation
  $p^*(\mathbf{d}_D|\mathbf{r}_B) = \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$
- Weighted Monte Carlo routine for sample based evaluation of posterior

Paper IV

JULLUM, M. & KOLBJØRNSEN, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics* **81**(3), R1–R13.

Properties

- Computationally cheap under stationarity conditions due to reuse of Gaussian approximations
- Offers a range of procedures with a trade-off between accuracy and computationally speed

Main thesis contributions

- Development of a principally new focused model selection strategy for selection among parametric and nonparametric type models
  - Few alternatives available
  - A new paradigm for the FIC
  - Beneficial theoretical behaviour
- Development of a new, locally focused procedure for Bayesian inversion within the geosciences
  - Combination of accuracy and computational speed seems to be out of reach for competing methodology

Main thesis contributions

- Development of a principally new focused model selection strategy for selection among parametric and nonparametric type models
  - Few alternatives available
  - A new paradigm for the FIC
  - Beneficial theoretical behaviour
- Development of a new, locally focused procedure for Bayesian inversion within the geosciences
  - Combination of accuracy and computational speed seems to be out of reach for competing methodology