Summary Statistics - CSML Reading Group

Alex Glyn-Davies

October 30, 2020

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$
- Simulated summaries $S(y)$ are then compared to observed $S(y_{\text{obs}})$ to accept/reject the sample

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$
- Simulated summaries $S(y)$ are then compared to observed $S(y_{\text{obs}})$ to accept/reject the sample
- Lower dimensional representation

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$
- Simulated summaries $S(y)$ are then compared to observed $S(y_{\text{obs}})$ to accept/reject the sample
- Lower dimensional representation $\rightarrow$ improved acceptance rate

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$
- Simulated summaries $S(y)$ are then compared to observed $S(y_{\text{obs}})$ to accept/reject the sample
- Lower dimensional representation $\rightarrow$ improved acceptance rate
- Optimal $S(y)$ would be "minimal sufficient" statistics

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$
- Simulated summaries $S(y)$ are then compared to observed $S(y_{\text{obs}})$ to accept/reject the sample
- Lower dimensional representation $\rightarrow$ improved acceptance rate
- Optimal $S(y)$ would be "minimal sufficient" statistics
- Often these are not available

Motivation

- When dealing with high-dimensional data y_{obs}, ABC algorithms use lower-dimensional summary statistics $S(y)$
- Simulated summaries $S(y)$ are then compared to observed $S(y_{\text{obs}})$ to accept/reject the sample
- Lower dimensional representation $\rightarrow$ improved acceptance rate
- Optimal $S(y)$ would be "minimal sufficient" statistics
- Often these are not available $\rightarrow$ resort to summary statistics

Sufficiency

- Want to infer parameters θ from the data y_{obs}

Sufficiency

- Want to infer parameters θ from the data y_{obs}
- Idea of sufficiency is to find statistics $S(y)$ of the data that summarise the information about θ

Sufficiency

- Want to infer parameters θ from the data y_{obs}
- Idea of sufficiency is to find statistics $S(y)$ of the data that summarise the information about θ

Definition (Bayes Sufficiency)

For any prior distribution of θ, the posterior density

$$f(\theta|y, S(y)) = f(\theta|S(y))$$

Sufficiency

Theorem (Fisher-Pitman-Koopman-Darmois)

With i.i.d. sampling from a model, exponential families are the only models for which there are sufficient statistics whose dimensions remain bounded as the sample size grows.

Sufficiency

Theorem (Fisher-Pitman-Koopman-Darmois)

With i.i.d. sampling from a model, exponential families are the only models for which there are sufficient statistics whose dimensions remain bounded as the sample size grows.

- This clearly presents a problem

Sufficiency

Theorem (Fisher-Pitman-Koopman-Darmois)

With i.i.d. sampling from a model, exponential families are the only models for which there are sufficient statistics whose dimensions remain bounded as the sample size grows.

- This clearly presents a problem
- Most of the time when dealing with exponential family model, we have access to a tractable likelihood

Sufficiency

Theorem (Fisher-Pitman-Koopman-Darmois)

With i.i.d. sampling from a model, exponential families are the only models for which there are sufficient statistics whose dimensions remain bounded as the sample size grows.

- This clearly presents a problem
- Most of the time when dealing with exponential family model, we have access to a tractable likelihood $\rightarrow$ ABC not typically required

Sufficiency

Theorem (Fisher-Pitman-Koopman-Darmois)

With i.i.d. sampling from a model, exponential families are the only models for which there are sufficient statistics whose dimensions remain bounded as the sample size grows.

- This clearly presents a problem
- Most of the time when dealing with exponential family model, we have access to a tractable likelihood $\rightarrow$ ABC not typically required
- Need methods for selecting appropriate low dimensional *insufficient* summaries

Summary Statistic Selection Methods

Summary Statistic Selection Methods

Choice of $S(y)$ will impact the **efficiency** and **accuracy** of ABC

Summary Statistic Selection Methods

Choice of $S(y)$ will impact the **efficiency** and **accuracy** of ABC

First two methods rely on training data and candidate summary statistics $z = (z_1, z_2, \ldots, z_k)$ where each z_i is a scalar function of data y

- Subset selection
- Projection methods

Summary Statistic Selection Methods

Choice of $S(y)$ will impact the **efficiency** and **accuracy** of ABC

First two methods rely on training data and candidate summary statistics $z = (z_1, z_2, \ldots, z_k)$ where each z_i is a scalar function of data y

- Subset selection
- Projection methods
- Auxiliary likelihood

Last method uses an approximating model to provide a more tractable "auxiliary" likelihood to derive summary statistics from

Summary Statistic Selection Methods

Choice of $S(y)$ will impact the **efficiency** and **accuracy** of ABC

First two methods rely on training data and candidate summary statistics $z = (z_1, z_2, \ldots, z_k)$ where each z_i is a scalar function of data y

- Subset selection
- Projection methods
- Auxiliary likelihood

Last method uses an approximating model to provide a more tractable "auxiliary" likelihood to derive summary statistics from

All these approaches require subjective input from the user

Example Data Features

Example data features used for Estimation of mutation rate in coalescent simulation (Nunes and Balding, 2010).

Table 1: The pool of summary statistics Ω considered for summarising datasets of DNA sequence haplotypes in the simulation study. For each statistic, we show the number of observed datasets (out of 100) for which it was included in the optimal set in univariate, unadjusted ABC inference by the methods described in the text.

		Selected for θ (%)			Selected for ρ (%)		
Statistic	Description	AS	ME	2-stage	AS	ME	2-stage
C_1	no. of segregating sites	75	67	100	73	67	97
C_2	Uniform[0,25] random variable	4	3	0	2	5	0
C_3	mean no. of differences over all pairs of haplotypes	27	54	25	52	30	19
C_4	25*(mean r^2 across pairs separated by $<$ 10% of the simulated genomic region)	56	35	50	35	59	78
C_5	no. of distinct haplotypes	43	19	20	78	73	100
C_6	frequency of the most common haplotype	36	20	1	11	23	2
C_7	no. of singleton haplotypes	16	14	5	16	31	5

Figure: Nunes and Balding, 2010

Example Data Features

Data features used for Random walk models (Barnes et al. 2012).

- S1 Mean square displacement.
- S2 Mean x and y displacement.
- S3 Mean square x and y displacement.
- S4 Straightness index.
- S5 Eigenvalues of gyration tensor (reference random walks book).

Applying our summary statistic selection framework to data simulated from the three different models over 100

Figure: Barnes et al.

Subset Selection Methods

- Attempts to find a subset of z that produces a low dimensional approximately sufficient set of statistics S'

Subset Selection Methods

- Attempts to find a subset of z that produces a low dimensional approximately sufficient set of statistics S'
- Requires training data, simplest way is to sample (θ, y) pairs by sampling θ from prior, then generating y

Subset Selection Methods

- Attempts to find a subset of z that produces a low dimensional approximately sufficient set of statistics S'
- Requires training data, simplest way is to sample (θ, y) pairs by sampling θ from prior, then generating y
- Other methods such as using pilot ABC run with $S(y) = z$, and using accepted simulations as training data

Subset Selection Methods

- Attempts to find a subset of z that produces a low dimensional approximately sufficient set of statistics S'
- Requires training data, simplest way is to sample (θ, y) pairs by sampling θ from prior, then generating y
- Other methods such as using pilot ABC run with $S(y) = z$, and using accepted simulations as training data
- Approximate sufficiency, Entropy minimisation, Mutual information maximisation

Subset Selection Methods

- Attempts to find a subset of z that produces a low dimensional approximately sufficient set of statistics S'
- Requires training data, simplest way is to sample (θ, y) pairs by sampling θ from prior, then generating y
- Other methods such as using pilot ABC run with $S(y) = z$, and using accepted simulations as training data
- Approximate sufficiency, Entropy minimisation, Mutual information maximisation
- Good for producing interpretable summaries - subset of interpretable candidates z more interpretable than some projection of z

Subset Selection Methods

- Attempts to find a subset of z that produces a low dimensional approximately sufficient set of statistics S'
- Requires training data, simplest way is to sample (θ, y) pairs by sampling θ from prior, then generating y
- Other methods such as using pilot ABC run with $S(y) = z$, and using accepted simulations as training data
- Approximate sufficiency, Entropy minimisation, Mutual information maximisation
- Good for producing interpretable summaries - subset of interpretable candidates z more interpretable than some projection of z
- Problems include large computational expense, often ABC must be run on all candidate subsets of z

Projection Methods

- Want to find a low dimensional **projection** of z

Projection Methods

- Want to find a low dimensional **projection** of z
- Partial Least Squares (PLS), Linear Regression, Boosting

Projection Methods

- Want to find a low dimensional **projection** of z
- Partial Least Squares (PLS), Linear Regression, Boosting
- Also requires training data

Projection Methods

- Want to find a low dimensional **projection** of z
- Partial Least Squares (PLS), Linear Regression, Boosting
- Also requires training data
- Less computationally expensive than subset selection methods

Projection Methods

- Want to find a low dimensional **projection** of z
- Partial Least Squares (PLS), Linear Regression, Boosting
- Also requires training data
- Less computationally expensive than subset selection methods
- Can use more candidate summaries because of this

Projection Methods

- Want to find a low dimensional **projection** of z
- Partial Least Squares (PLS), Linear Regression, Boosting
- Also requires training data
- Less computationally expensive than subset selection methods
- Can use more candidate summaries because of this
- Wider search space of candidate summaries - linear/non-linear combinations, not just subsets

Projection Methods

- Want to find a low dimensional **projection** of z
- Partial Least Squares (PLS), Linear Regression, Boosting
- Also requires training data
- Less computationally expensive than subset selection methods
- Can use more candidate summaries because of this
- Wider search space of candidate summaries - linear/non-linear combinations, not just subsets
- All methods apply to multi-dimensional θ

Auxiliary Likelihood

- Need an **approximate** and **tractable** likelihood for the data

Auxiliary Likelihood

- Need an **approximate** and **tractable** likelihood for the data
- Auxiliary likelihood $p_A(y|\phi)$, auxiliary parameters ϕ don't need to correspond to generative parameters θ

Auxiliary Likelihood

- Need an **approximate** and **tractable** likelihood for the data
- Auxiliary likelihood $p_A(y|\phi)$, auxiliary parameters ϕ don't need to correspond to generative parameters θ
- Maximum likelihood estimators, Likelihood distance, Scores

Auxiliary Likelihood

- Need an **approximate** and **tractable** likelihood for the data
- Auxiliary likelihood $p_A(y|\phi)$, auxiliary parameters ϕ don't need to correspond to generative parameters θ
- Maximum likelihood estimators, Likelihood distance, Scores
- No need for training data

Auxiliary Likelihood

- Need an **approximate** and **tractable** likelihood for the data
- Auxiliary likelihood $p_A(y|\phi)$, auxiliary parameters ϕ don't need to correspond to generative parameters θ
- Maximum likelihood estimators, Likelihood distance, Scores
- No need for training data
- Subjective choice of this approximating model - may be difficult/poorly approximating

Choice of Auxiliary Model

- General model with tractable likelihood e.g. Gaussian Mixture

Choice of Auxiliary Model

- General model with tractable likelihood e.g. Gaussian Mixture
- Approximate the generative likelihood with tractable alternative e.g. Composite likelihood

Choice of Auxiliary Model

- General model with tractable likelihood e.g. Gaussian Mixture
- Approximate the generative likelihood with tractable alternative e.g. Composite likelihood
- Want small number of parameters to produce low-dimensional summaries

Choice of Auxiliary Model

- General model with tractable likelihood e.g. Gaussian Mixture
- Approximate the generative likelihood with tractable alternative e.g. Composite likelihood
- Want small number of parameters to produce low-dimensional summaries
- Hard to assess whether the auxiliary likelihood is producing informative summaries for the generative model

Choice of Auxiliary Model

- General model with tractable likelihood e.g. Gaussian Mixture
- Approximate the generative likelihood with tractable alternative e.g. Composite likelihood
- Want small number of parameters to produce low-dimensional summaries
- Hard to assess whether the auxiliary likelihood is producing informative summaries for the generative model

Example

4.1. The structural model: An Ornstein-Uhlenbeck type stochastic volatility model

Our structural model $\mathcal{M}_S$ is defined in terms of the following two stochastic differential equations:

$$\mathrm{d}\, x^*(t) = (\mu + \beta\sigma^2(t))\,\mathrm{d}\, t + \sigma(t)\,\mathrm{d}\, W(t) \tag{4.1}$$

$$\mathrm{d}\, \sigma^2(t) = -\lambda\sigma^2(t)\,\mathrm{d}\, t + \mathrm{d}\, Z(\lambda t). \tag{4.2}$$

Here we denote with $(x^*(t))_{t\geq 0}$ the log price process of an asset, $(W(t))_{t\geq 0}$ is a standard Brownian motion and $(\sigma^2(t))_{t\geq 0}$ is the underlying latent *instantaneous volatility process* of OU type, independent of $(W(t))_{t\geq 0}$, with $(Z(\lambda t))_{t\geq 0}$ being the *background driving Lévy process*

Discussion

- Blum et al. (2013) - ‘What is very apparent from this study is that there is no single “best” method of dimension reduction for ABC.’

Discussion

- Blum et al. (2013) - 'What is very apparent from this study is that there is no single "best" method of dimension reduction for ABC.'
- With low $k = \dim(z)$, subset selection methods are computationally feasible and perform best

Discussion

- Blum et al. (2013) - 'What is very apparent from this study is that there is no single "best" method of dimension reduction for ABC.'
- With low $k = \dim(z)$, subset selection methods are computationally feasible and perform best
- With high k, projection methods are favoured

Discussion

- Blum et al. (2013) - 'What is very apparent from this study is that there is no single "best" method of dimension reduction for ABC.'
- With low $k = \dim(z)$, subset selection methods are computationally feasible and perform best
- With high k, projection methods are favoured
- Chapter could have discussed the selection of data features z - could initially better selection of z reduce the need for complex and expensive summary selection?

Discussion

- Blum et al. (2013) - 'What is very apparent from this study is that there is no single "best" method of dimension reduction for ABC.'
- With low $k = \dim(z)$, subset selection methods are computationally feasible and perform best
- With high k, projection methods are favoured
- Chapter could have discussed the selection of data features z - could initially better selection of z reduce the need for complex and expensive summary selection?
- Is there any room for improving this initial selection of features?

Discussion

- Blum et al. (2013) - 'What is very apparent from this study is that there is no single "best" method of dimension reduction for ABC.'
- With low $k = \dim(z)$, subset selection methods are computationally feasible and perform best
- With high k, projection methods are favoured
- Chapter could have discussed the selection of data features z - could initially better selection of z reduce the need for complex and expensive summary selection?
- Is there any room for improving this initial selection of features?
- No analysis of how these methods affect the accuracy of the posterior approximation

Other Approaches

- **K2-ABC: Approximate Bayesian Computation with Kernel Embeddings** - Park et al., (2016)

Other Approaches

- **K2-ABC: Approximate Bayesian Computation with Kernel Embeddings** - Park et al., (2016)

  - Circumvents need for selecting summary statistics - uses MMD to give a dissimilarity measure between y_{obs} and simulated y

Other Approaches

- **K2-ABC: Approximate Bayesian Computation with Kernel Embeddings** - Park et al., (2016)

  - Circumvents need for selecting summary statistics - uses MMD to give a dissimilarity measure between y_{obs} and simulated y

  - Does need training data to learn the regression in RKHS

Other Approaches

- **K2-ABC: Approximate Bayesian Computation with Kernel Embeddings** - Park et al., (2016)

  - Circumvents need for selecting summary statistics - uses MMD to give a dissimilarity measure between y_{obs} and simulated y

  - Does need training data to learn the regression in RKHS

  - Still need to pick the characteristic kernel, this is subjective

Other Approaches

- **K2-ABC: Approximate Bayesian Computation with Kernel Embeddings** - Park et al., (2016)

  - Circumvents need for selecting summary statistics - uses MMD to give a dissimilarity measure between y_{obs} and simulated y

  - Does need training data to learn the regression in RKHS

  - Still need to pick the characteristic kernel, this is subjective

- **Approximate Bayesian computation via the energy statistic** - Nguyen et al., (2020)

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

$$\left| \frac{\hat{\pi}_{\text{ABC}}(\theta | S'(y_{\text{obs}}))}{\hat{\pi}_{\text{ABC}}(\theta | S(y_{\text{obs}}))} - 1 \right| > T(\theta)$$

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

$$\left| \frac{\hat{\pi}_{\mathrm{ABC}}(\theta | S'(y_{\mathrm{obs}}))}{\hat{\pi}_{\mathrm{ABC}}(\theta | S(y_{\mathrm{obs}}))} - 1 \right| > T(\theta)$$

  Motivated by the definition of Bayesian sufficiency

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

$$\left| \frac{\hat{\pi}_{\mathrm{ABC}}(\theta | S'(y_{\mathrm{obs}}))}{\hat{\pi}_{\mathrm{ABC}}(\theta | S(y_{\mathrm{obs}}))} - 1 \right| > T(\theta)$$

  Motivated by the definition of Bayesian sufficiency
  Only for scalar θ

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

  $$\left|\frac{\hat{\pi}_{\mathrm{ABC}}(\theta|S'(y_{\mathrm{obs}}))}{\hat{\pi}_{\mathrm{ABC}}(\theta|S(y_{\mathrm{obs}}))} - 1\right| > T(\theta)$$

  Motivated by the definition of Bayesian sufficiency
  Only for scalar θ
- **Entropy Minimisation** - Nunes and Balding, 2010

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

  $$\left| \frac{\hat{\pi}_{\mathrm{ABC}}(\theta | S'(y_{\mathrm{obs}}))}{\hat{\pi}_{\mathrm{ABC}}(\theta | S(y_{\mathrm{obs}}))} - 1 \right| > T(\theta)$$

  Motivated by the definition of Bayesian sufficiency
  Only for scalar θ
- **Entropy Minimisation** - Nunes and Balding, 2010
  Complex two-step approach.

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

$$\left| \frac{\hat{\pi}_{\text{ABC}}(\theta | S'(y_{\text{obs}}))}{\hat{\pi}_{\text{ABC}}(\theta | S(y_{\text{obs}}))} - 1 \right| > T(\theta)$$

  Motivated by the definition of Bayesian sufficiency
  Only for scalar θ

- **Entropy Minimisation** - Nunes and Balding, 2010
  Complex two-step approach.
  - Minimise estimate of ABC posterior entropy to pick S_{ME}, retain n_{obs} 'best' datasets for training

Subset Selection Methods

- **Approximate Sufficiency** - Joyce and Marjoram, 2008
  Candidate statistics randomly added, and are only accepted if there is a great enough change in posterior approximation

  $$\left|\frac{\hat{\pi}_{\mathrm{ABC}}(\theta|S'(y_{\mathrm{obs}}))}{\hat{\pi}_{\mathrm{ABC}}(\theta|S(y_{\mathrm{obs}}))} - 1\right| > T(\theta)$$

  Motivated by the definition of Bayesian sufficiency
  Only for scalar θ
- **Entropy Minimisation** - Nunes and Balding, 2010
  Complex two-step approach.
  - Minimise estimate of ABC posterior entropy to pick S_{ME}, retain n_{obs} 'best' datasets for training
  - Repeatedly run rejection-ABC, minimise the RMSE of parameters compared to best datasets over subsets of z

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

  - Equivalent to minimising the expected entropy over y

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

  - Equivalent to minimising the expected entropy over y

  Admits dim $\theta > 1$

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

  - Equivalent to minimising the expected entropy over y

  Admits dim $\theta > 1$
- **Regularisation**

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

  - Equivalent to minimising the expected entropy over y

  Admits dim $\theta > 1$

- **Regularisation**
  - Local-linear regression model with response θ, and covariates z, in the region of $S(y_{\text{obs}})$

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

  - Equivalent to minimising the expected entropy over y

  Admits dim $\theta > 1$

- **Regularisation**
  - Local-linear regression model with response θ, and covariates z, in the region of $S(y_{\text{obs}})$

  - Use the AIC/BIC criterion to penalise complexity, and select relevant data features

Subset Selection Methods

- **Mutual Information Maximisation** - Barnes et al., 2012
  $I(\theta; S(y)) = I(\theta; y)$ iif $S(y)$ sufficient

  - Add in z_i that maximises estimated $\mathcal{KL}$-divergence between ABC posteriors

  - Equivalent to minimising the expected entropy over y

  Admits dim $\theta > 1$

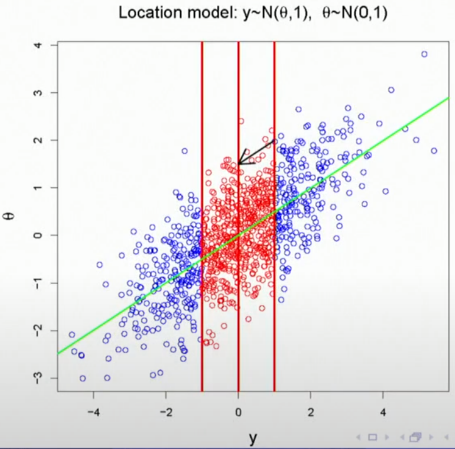
- **Regularisation**
  - Local-linear regression model with response θ, and covariates z, in the region of $S(y_{\text{obs}})$

  - Use the AIC/BIC criterion to penalise complexity, and select relevant data features

  - Post-processing allows for samples $\|S(y_{\text{obs}}) - S(y)\| < h$ to be adjusted based on the local-linear regression

Further refinements - regression

Beaumont, Zhang and Balding, 2002, Blum, 2010, Blum and François, 2010



Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \mathrm{Cov}(u_i, \theta_j)^2$, s.t. $\mathrm{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \mathrm{Cov}(u_i, \theta_j)^2$, s.t. $\mathrm{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

  - Produces linear combinations of z that have high covariance with θ, and are uncorrelated with each other

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \mathrm{Cov}(u_i, \theta_j)^2$, s.t. $\mathrm{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

  - Produces linear combinations of z that have high covariance with θ, and are uncorrelated with each other

  - Pick the c first components as summary statistics, to reduce the dimension (they use a cross-validation procedure to select c)

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \mathrm{Cov}(u_i, \theta_j)^2$, s.t. $\mathrm{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

  - Produces linear combinations of z that have high covariance with θ, and are uncorrelated with each other

  - Pick the c first components as summary statistics, to reduce the dimension (they use a cross-validation procedure to select c)

- **Linear Regression** - Fearnhead and Prangle, 2012

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \mathrm{Cov}(u_i, \theta_j)^2$, s.t. $\mathrm{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

  - Produces linear combinations of z that have high covariance with θ, and are uncorrelated with each other

  - Pick the c first components as summary statistics, to reduce the dimension (they use a cross-validation procedure to select c)

- **Linear Regression** - Fearnhead and Prangle, 2012

  - Fit linear model to training data, $\theta \sim \mathcal{N}(Az + b, \Sigma)$

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \text{Cov}(u_i, \theta_j)^2$, s.t. $\text{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

  - Produces linear combinations of z that have high covariance with θ, and are uncorrelated with each other

  - Pick the c first components as summary statistics, to reduce the dimension (they use a cross-validation procedure to select c)

- **Linear Regression** - Fearnhead and Prangle, 2012

  - Fit linear model to training data, $\theta \sim \mathcal{N}(Az + b, \Sigma)$

  - Motivated by $S(y) = \mathbb{E}[\theta|y]$ being optimal choice of S to minimise quadratic loss of parameter means in the target distribution $\pi(\theta|S(y_{\text{obs}}))$ when $h = 0$

Projection Methods

- **Partial Least Squares** - Wegmann et al., 2009

  - ith PLS component $u_i = \alpha_i^T z$ maximises $\sum_{j=1}^{p} \text{Cov}(u_i, \theta_j)^2$, s.t. $\text{Cov}(u_i, u_j) = 0$ for $j < i$. Also normalisation constraint $\alpha_i^T \alpha_i = 1$

  - Produces linear combinations of z that have high covariance with θ, and are uncorrelated with each other

  - Pick the c first components as summary statistics, to reduce the dimension (they use a cross-validation procedure to select c)

- **Linear Regression** - Fearnhead and Prangle, 2012

  - Fit linear model to training data, $\theta \sim \mathcal{N}(Az + b, \Sigma)$

  - Motivated by $S(y) = \mathbb{E}[\theta|y]$ being optimal choice of S to minimise quadratic loss of parameter means in the target distribution $\pi(\theta|S(y_{\text{obs}}))$ when $h = 0$

Projection Methods

- **Boosting** - Aeschbacher et al., 2012

Projection Methods

- **Boosting** - Aeschbacher et al., 2012

  - Non-linear regression method, uses training data and outputs predictors $\hat{\theta}(y)$ of $\mathbb{E}(\theta|y)$, which are used as summary statistics

Projection Methods

- **Boosting** - Aeschbacher et al., 2012

  - Non-linear regression method, uses training data and outputs predictors $\hat{\theta}(y)$ of $\mathbb{E}(\theta|y)$, which are used as summary statistics

  - Generates an ensemble of weak learners to construct a strong learner