Inference under discrepancy

Richard Wilkinson

University of Sheffield

Inference under discrepancy

How should we do inference if the model is imperfect?

[1] Even if we can't agree about it!

Inference under discrepancy

How should we do inference if the model is imperfect?

Data generating process

$$y \sim G$$

[1] Even if we can't agree about it!

Inference under discrepancy

How should we do inference if the model is imperfect?
Data generating process

$$y \sim G$$

Model (complex simulator, finite dimensional parameter)

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

If $G = F_{\theta_0} \in \mathcal{F}$ then we know what to do[1].

[1] Even if we can't agree about it!

Inference under discrepancy

How should we do inference if the model is imperfect?
Data generating process

$$y \sim G$$

Model (complex simulator, finite dimensional parameter)

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

If $G = F_{\theta_0} \in \mathcal{F}$ then we know what to do[1].

How should we proceed if

$$G \notin \mathcal{F}$$

[1] Even if we can't agree about it!

Inference under discrepancy

How should we do inference if the model is imperfect?
Data generating process

$$y \sim G$$

Model (complex simulator, finite dimensional parameter)

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

If $G = F_{\theta_0} \in \mathcal{F}$ then we know what to do[1].

How should we proceed if

$$G \notin \mathcal{F}$$

Interest lies in inference of θ not calibrated prediction.

[1]Even if we can't agree about it!

An appealing idea

Kennedy an O'Hagan 2001

Lets acknowledge that most models are imperfect.

An appealing idea

Kennedy an O'Hagan 2001

Lets acknowledge that most models are imperfect.

Can we expand the class of models by adding a Gaussian process (GP) to our simulator?

If $f_\theta(x)$ is our simulator, y the observation, then perhaps we can correct f by modelling

$$y = f_{\theta^*}(x) + \delta(x) \quad \text{where} \quad \delta \sim GP$$

An appealing idea

Kennedy an O'Hagan 2001

Lets acknowledge that most models are imperfect.

Can we expand the class of models by adding a Gaussian process (GP) to our simulator?

If $f_\theta(x)$ is our simulator, y the observation, then perhaps we can correct f by modelling

$$y = f_{\theta^*}(x) + \delta(x) \quad \text{where} \quad \delta \sim GP$$

This greatly expands $\mathcal{F}$ into a non-parametric world.

An appealing, but flawed, idea

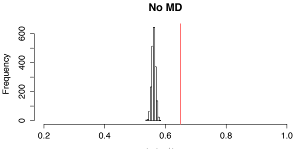
Kennedy and O'Hagan 2001, Brynjarsdottir and O'Hagan 2014

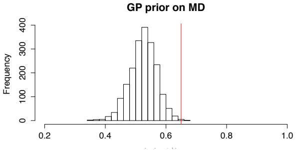
Simulator

$$f_\theta(x) = \theta x$$

Reality

$$g(x) = \frac{\theta x}{1 + \frac{x}{a}} \quad \theta = 0.65, a = 20$$





Bolting on a GP can correct your predictions, but won't necessarily fix your inference.

Dangers of non-parametric model extensions

There are (at least) two problems with this approach:

- We may still find $G \notin \mathcal{F}$
- Identifiability

Dangers of non-parametric model extensions

There are (at least) two problems with this approach:

- We may still find $G \notin \mathcal{F}$
- Identifiability
  - A GP is an incredibly complex infinite dimensional model, which is not necessarily identified even asymptotically. The posterior can concentrate not on a point, but on some sub manifold of parameter space, and the projection of the prior on this space continues to impact the posterior even as more and more data are collected.

    ie We never forget the prior, but the prior is to complex to understand

Dangers of non-parametric model extensions

There are (at least) two problems with this approach:

- We may still find $G \notin \mathcal{F}$
- Identifiability
  - A GP is an incredibly complex infinite dimensional model, which is not necessarily identified even asymptotically. The posterior can concentrate not on a point, but on some sub manifold of parameter space, and the projection of the prior on this space continues to impact the posterior even as more and more data are collected.

    ie We never forget the prior, but the prior is to complex to understand
  - Brynjarsdottir and O'Hagan 2014 try to model their way out of trouble with prior information - which is great if you have it.

Dangers of non-parametric model extensions

There are (at least) two problems with this approach:

- We may still find $G \notin \mathcal{F}$
- Identifiability
  - A GP is an incredibly complex infinite dimensional model, which is not necessarily identified even asymptotically. The posterior can concentrate not on a point, but on some sub manifold of parameter space, and the projection of the prior on this space continues to impact the posterior even as more and more data are collected.

    ie We never forget the prior, but the prior is to complex to understand
  - Brynjarsdottir and O'Hagan 2014 try to model their way out of trouble with prior information - which is great if you have it.
  - Wong et al 2017 impose identifiability (for δ and θ) by giving up and identifying

$$\theta^* = \arg\min_{\theta} \int (\zeta(x) - f_\theta(x))^2 d\pi(x)$$

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 2, *pp.* 635–648

A frequentist approach to computer model calibration

Inferential approaches

- Maximum likelihood/minimum-distance
- Bayes(ish)
- History matching (HM)/ABC type methods (thresholding)

2

Inferential approaches

- Maximum likelihood/minimum-distance
- Bayes(ish)
- History matching (HM)/ABC type methods (thresholding)

We'll consider how they behave for well-specified and mis-specified models.

Try to understand why (at least anecdotally) HM and ABC seem to work well in mis-specified cases.

2

Inferential approaches

- Maximum likelihood/minimum-distance
- Bayes(ish)
- History matching (HM)/ABC type methods (thresholding)

We'll consider how they behave for well-specified and mis-specified models.

Try to understand why (at least anecdotally) HM and ABC seem to work well in mis-specified cases.

Big question[2] is what properties would we like our inferential approach to possess.

[2] To which I have no answer

Maximum likelihood

Maximum likelihood estimator

$$\hat{\theta}_n = \arg\max_{\theta} l(y|\theta)$$

If $G = F_{\theta_0} \in \mathcal{F}$, then (under some conditions)

$$\hat{\theta}_n \to \theta_0 \text{ almost surely as } \quad n \to \infty$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\Rightarrow} N(0, \mathcal{I}^{-1}(\theta_0))$$

Asymptotic consistency, efficiency, normality.

Maximum likelihood

Maximum likelihood estimator

$$\hat{\theta}_n = \arg\max_{\theta} l(y|\theta)$$

If $G = F_{\theta_0} \in \mathcal{F}$, then (under some conditions)

$$\hat{\theta}_n \to \theta_0 \text{ almost surely as } \quad n \to \infty$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\Rightarrow} N(0, \mathcal{I}^{-1}(\theta_0))$$

Asymptotic consistency, efficiency, normality.
If $G \notin \mathcal{F}$

$$\hat{\theta}_n \to \theta^* = \arg\min_{\theta} D_{KL}(G, F_\theta) \text{ almost surely}$$

$$= \arg\min_{\theta} \int \log \frac{dG}{dF_\theta} dG$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\Rightarrow} N(0, V^{-1})$$

Bayes

Bayesian posterior

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta)$$

If $G = F_{\theta_0} \in \mathcal{F}$

$$\pi(\theta|y) \overset{d}{\Rightarrow} N(\theta_0, \mathcal{I}^{-1}(\theta_0)) \text{ as } \quad n \to \infty$$

Bernstein-von Mises theorem: we forget the prior, and get asymptotic concentration and normality.
This also requires (a long list of) identifiability conditions to hold.

Bayes

Bayesian posterior

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta)$$

If $G = F_{\theta_0} \in \mathcal{F}$

$$\pi(\theta|y) \overset{d}{\Rightarrow} N(\theta_0, \mathcal{I}^{-1}(\theta_0)) \text{ as } \quad n \to \infty$$

Bernstein-von Mises theorem: we forget the prior, and get asymptotic concentration and normality.
This also requires (a long list of) identifiability conditions to hold.

If $G \notin \mathcal{F}$, we still get asymptotic concentration (and possibly normality) but to θ^* (the pseudo-true value).

> *"there is no obvious meaning for Bayesian analysis in this case"*

Bayes

Bayesian posterior

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta)$$

If $G = F_{\theta_0} \in \mathcal{F}$

$$\pi(\theta|y) \stackrel{d}{\Rightarrow} N(\theta_0, \mathcal{I}^{-1}(\theta_0)) \text{ as } \quad n \to \infty$$

Bernstein-von Mises theorem: we forget the prior, and get asymptotic concentration and normality.
This also requires (a long list of) identifiability conditions to hold.

If $G \notin \mathcal{F}$, we still get asymptotic concentration (and possibly normality) but to θ^* (the pseudo-true value).

> *"there is no obvious meaning for Bayesian analysis in this case"*

Often with non-parametric models (eg GPs), we don't even get this convergence to the pseudo-true value due to lack of identifiability.

History matching and ABC

History matching seeks to find a NROY set

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_{\theta,y}) \leq 3\}$$

where

$$S_{HM}(F_\theta) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\mathbb{V}\text{ar}_{F_\theta}(Y)}}$$

History matching and ABC

History matching seeks to find a NROY set

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_{\theta,y}) \leq 3\}$$

where

$$S_{HM}(F_\theta) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\mathbb{V}\text{ar}_{F_\theta}(Y)}}$$

ABC approximates the posterior as

$$\pi_\epsilon(\theta) \propto \pi(\theta)\mathbb{E}(\mathbb{I}_{S(\hat{F}_\theta,y)\leq\epsilon})$$

for some choice of S (typically $S(\hat{F}_\theta, y) = \rho(\eta(y), \eta(y'))$ where $y' \sim F_\theta$) and ϵ.

History matching and ABC

History matching seeks to find a NROY set

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_{\theta,y}) \leq 3\}$$

where

$$S_{HM}(F_\theta) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\mathbb{V}\text{ar}_{F_\theta}(Y)}}$$

ABC approximates the posterior as

$$\pi_\epsilon(\theta) \propto \pi(\theta)\mathbb{E}(\mathbb{I}_{S(\hat{F}_\theta, y) \leq \epsilon})$$

for some choice of S (typically $S(\hat{F}_\theta, y) = \rho(\eta(y), \eta(y'))$ where $y' \sim F_\theta$) and ϵ.

They have thresholding of a score in common and are algorithmically comparable.

History matching and ABC

These methods (anecdotally) seem to work better in mis-specified situations.

Why?

History matching and ABC

These methods (anecdotally) seem to work better in mis-specified situations.

Why?

They differ from likelihood based approaches in that

- They only use some aspect of the simulator output
  - Typically we hand pick which simulator outputs to compare, and weight them on a case by case basis.
- Potentially use generalised scores/loss-functions
- The thresholding type nature potentially makes them somewhat conservative

History matching and ABC

These methods (anecdotally) seem to work better in mis-specified situations.

Why?

They differ from likelihood based approaches in that

- They only use some aspect of the simulator output
  - Typically we hand pick which simulator outputs to compare, and weight them on a case by case basis.
- Potentially use generalised scores/loss-functions
- The thresholding type nature potentially makes them somewhat conservative

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.
- Asymptotic concentration or normality?

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.
- ~~Asymptotic concentration or normality?~~

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.
- ~~Asymptotic concentration or normality?~~
- Frequency properties?

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.
- ~~Asymptotic concentration or normality?~~
- Frequency properties?
  - I wouldn't object but seems impossible for subjective priors.

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.
- ~~Asymptotic concentration or normality?~~
- Frequency properties?
  - I wouldn't object but seems impossible for subjective priors.
- Coherence?

What properties do we want?

Do any of these approaches have favourable properties/characteristics for inference under discrepancy? Particularly when the discrepancy model is crude?

- Consistency?
  - I don't want inconsistency.
- ~~Asymptotic concentration or normality?~~
- Frequency properties?
  - I wouldn't object but seems impossible for subjective priors.
- Coherence?
- Robustness to small mis-specifications?

Generalized scores

Likelihood based methods are notoriously sensitive to mis-specification.

Generalized scores

Likelihood based methods are notoriously sensitive to mis-specification. Consider scoring rules instead. If we forecast F, observe y, then we receive score

$$S(F, y)$$

Generalized scores

Likelihood based methods are notoriously sensitive to mis-specification. Consider scoring rules instead. If we forecast F, observe y, then we receive score

$$S(F, y)$$

S is a proper score if

$$G = \arg\min_F \mathbb{E}_{Y\sim G} S(F, Y)$$

i.e. predicting G gives the best possibly score.

- Encourages honest reporting

Generalized scores

Likelihood based methods are notoriously sensitive to mis-specification. Consider scoring rules instead. If we forecast F, observe y, then we receive score

$$S(F, y)$$

S is a proper score if

$$G = \arg\min_F \mathbb{E}_{Y \sim G} S(F, Y)$$

i.e. predicting G gives the best possibly score.

- Encourages honest reporting

Examples:

- Log-likelihood $S(F, y) = -\log f(y)$
- Tsallis-score $(\gamma - 1) \int f(x)^\alpha dx - \gamma f(y)^{\alpha - 1}$

Minimum scoring rule estimation (Dawid *et al.* 2014 etc) uses

$$\hat{\theta} = \arg\min_{\theta} S(F_{\theta}, y)$$

Minimum scoring rule estimation (Dawid *et al.* 2014 etc) uses

$$\hat{\theta} = \arg\min_{\theta} S(F_\theta, y)$$

For proper scores

$$\mathbb{E}_{\theta_0}\left(\frac{\partial}{\partial\theta} S(F_\theta, y)\Big|_{\theta=\theta_0}\right) = \frac{\partial}{\partial\theta}\mathbb{E}_{\theta_0} S(F_\theta, y)\Big|_{\theta=\theta_0}$$
$$= 0$$

so we have an unbiased estimating equation, and hence get asymptotic consistency for well-specified models. We also get asymptotic normality.

Dawid *et al.* 2014 show that if

- $\nabla_\theta f_\theta(x)$ is bounded in x for all θ
- Bregman gauge of scoring rule is locally bounded

then the minimum scoring rule estimator $\hat{\theta}$ is B-robust

- i.e. it has bounded influence function

$$IF(x; \hat{\theta}, F_\theta) = \lim_{\epsilon \to 0} \frac{\hat{\theta}(\epsilon\delta_x + (1-\epsilon)F_\theta) - \hat{\theta}(F_\theta)}{\epsilon}$$

  i.e. if F_θ is infected by outlier at x, this doesn't unduly affect the inference.

Note both ABC and HM are B-robust in this sense, but using the log-likelihood is not.

Dawid *et al.* 2014 show that if

- $\nabla_\theta f_\theta(x)$ is bounded in x for all θ
- Bregman gauge of scoring rule is locally bounded

then the minimum scoring rule estimator $\hat{\theta}$ is B-robust

- i.e. it has bounded influence function

$$IF(x; \hat{\theta}, F_\theta) = \lim_{\epsilon \to 0} \frac{\hat{\theta}(\epsilon\delta_x + (1-\epsilon)F_\theta) - \hat{\theta}(F_\theta)}{\epsilon}$$

  i.e. if F_θ is infected by outlier at x, this doesn't unduly affect the inference.

Note both ABC and HM are B-robust in this sense, but using the log-likelihood is not.

What type of robustness do we want here?

Dawid *et al.* 2014 show that if

- $\nabla_\theta f_\theta(x)$ is bounded in x for all θ
- Bregman gauge of scoring rule is locally bounded

then the minimum scoring rule estimator $\hat{\theta}$ is B-robust

- i.e. it has bounded influence function

$$IF(x; \hat{\theta}, F_\theta) = \lim_{\epsilon \to 0} \frac{\hat{\theta}(\epsilon\delta_x + (1-\epsilon)F_\theta) - \hat{\theta}(F_\theta)}{\epsilon}$$

  i.e. if F_θ is infected by outlier at x, this doesn't unduly affect the inference.

Note both ABC and HM are B-robust in this sense, but using the log-likelihood is not.

What type of robustness do we want here?

Bayes like approaches

What about Bayesian like approaches with generalized scores?

Bayes like approaches

What about Bayesian like approaches with generalized scores?

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2016)
78, *Part* 5, *pp.* 1103–1130

A general framework for updating belief distributions

Bissiri et al. 2016 consider updating prior beliefs when parameter θ is connected to observations via a loss function $L(\theta, y)$.

Bayes like approaches

What about Bayesian like approaches with generalized scores?

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2016)
78, *Part* 5, *pp.* 1103–1130

A general framework for updating belief distributions

Bissiri et al. 2016 consider updating prior beliefs when parameter θ is connected to observations via a loss function $L(\theta, y)$.
They argue the update must be of the form

$$\pi(\theta|x) \propto \exp(-L(\theta, x))\pi(\theta)$$

via coherency arguments.
Note using log-likelihood as the loss function $(L(\theta, x) = -\log f_\theta(x))$ recovers Bayes.

Bayes like approaches

What about Bayesian like approaches with generalized scores?

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2016)
78, *Part* 5, *pp.* 1103–1130

A general framework for updating belief distributions

Bissiri et al. 2016 consider updating prior beliefs when parameter θ is connected to observations via a loss function $L(\theta, y)$.
They argue the update must be of the form

$$\pi(\theta|x) \propto \exp(-L(\theta, x))\pi(\theta)$$

via coherency arguments.
Note using log-likelihood as the loss function $(L(\theta, x) = -\log f_\theta(x))$ recovers Bayes.

Advantages of this include

- Allows focus solely on the quantities of interest.
  - Full Bayesian inference requires us to model the complete data distribution even when we're only interested in a low-dimensional summary statistic of the population.
- Deals better with mis-specification

Advantages of this include

- Allows focus solely on the quantities of interest.
  - Full Bayesian inference requires us to model the complete data distribution even when we're only interested in a low-dimensional summary statistic of the population.
- Deals better with mis-specification

Presumably the posterior may inherit some form of robustness from certain choices for the loss function, e.g., the bounded robust proper scores of Dawid *et al.* .

Advantages of this include

- Allows focus solely on the quantities of interest.
    - Full Bayesian inference requires us to model the complete data distribution even when we're only interested in a low-dimensional summary statistic of the population.
- Deals better with mis-specification

Presumably the posterior may inherit some form of robustness from certain choices for the loss function, e.g., the bounded robust proper scores of Dawid *et al.* .

Relates to the Bayes linear approach of Goldstein and Wooff which is also motivated by difficulties with specifying a complete model for the data.

HM and ABC thresholding

History matching was an approach designed for inference for mis-specified models.

$$S_{HM}(F_\theta) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\mathbb{V}\text{ar}_{F_\theta}(y)}}$$

Often applied in a Bayes linear type setting, with $\mathbb{V}\text{ar}_{F_\theta}(y)$ broken down into constituent parts

$$\mathbb{V}\text{ar}_{F_\theta}(y) = \mathbb{V}\text{ar}_{sim} + \mathbb{V}\text{ar}_{discrep} + \mathbb{V}\text{ar}_{emulator}$$

HM and ABC thresholding

History matching was an approach designed for inference for mis-specified models.

$$S_{HM}(F_\theta) = \frac{|\mathbb{E}_{F_\theta}(Y) - y|}{\sqrt{\mathbb{V}\mathrm{ar}_{F_\theta}(y)}}$$

Often applied in a Bayes linear type setting, with $\mathbb{V}\mathrm{ar}_{F_\theta}(y)$ broken down into constituent parts

$$\mathbb{V}\mathrm{ar}_{F_\theta}(y) = \mathbb{V}\mathrm{ar}_{sim} + \mathbb{V}\mathrm{ar}_{discrep} + \mathbb{V}\mathrm{ar}_{emulator}$$

Combined with the thresholding nature

$$\mathcal{P}_\theta = \{\theta : S_{HM}(\hat{F}_{\theta,y}) \leq 3\}$$

means we don't get asymptotic concentration.

- ABC shares similar properties if ϵ fixed at something reasonable.

$$\pi_\epsilon(\theta) \propto \pi(\theta)\mathbb{I}_{S(\hat{F}_\theta, y) \leq \epsilon}$$

  The indicator functions acts to add a ball of radius ϵ around the data, so that we only need to get within it.
- ϵ plays the same role as $\mathbb{V}\text{ar}_{discrep}$ in HM.

- ABC shares similar properties if ϵ fixed at something reasonable.

$$\pi_\epsilon(\theta) \propto \pi(\theta)\mathbb{I}_{S(\hat{F}_\theta, y)\leq\epsilon}$$

  The indicator functions acts to add a ball of radius ϵ around the data, so that we only need to get within it.
- ϵ plays the same role as $\mathbb{V}\text{ar}_{discrep}$ in HM.

Both approaches also allow the user to focus on aspects/summaries of the simulator output that either are of interest, or for which we believe the simulator is better specified.

- We discard information by only using some aspects of the simulator output, but perhaps to benefit of the inference

- ABC shares similar properties if ϵ fixed at something reasonable.

$$\pi_\epsilon(\theta) \propto \pi(\theta)\mathbb{I}_{S(\hat{F}_\theta, y) \leq \epsilon}$$

  The indicator functions acts to add a ball of radius ϵ around the data, so that we only need to get within it.
- ϵ plays the same role as $\mathbb{V}\mathrm{ar}_{discrep}$ in HM.

Both approaches also allow the user to focus on aspects/summaries of the simulator output that either are of interest, or for which we believe the simulator is better specified.

- We discard information by only using some aspects of the simulator output, but perhaps to benefit of the inference

Also

- Allow for crude/simple discrepancy characterization.
- Some form of robustness arises from the scores used.

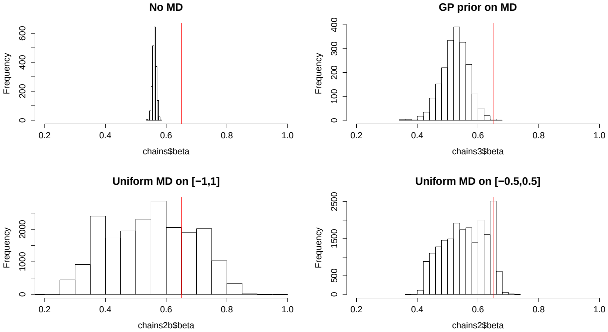
Brynjarsdottir *et al.* revisited

Simulator

$$f_\theta(x) = \theta x$$

Reality

$$g(x) = \frac{\theta x}{1 + \frac{x}{a}} \qquad \theta = 0.65, a = 20$$



Recent work in ABC

Recent work on ABC has sought to move away from the use of summaries

- Bernton *et al.* 2017 look at Bayes like procedures based on the Wasserstein distance (get different pseudo-true value)
- Park *et al.* 2015 look at using kernel mean embeddings of distributions to also avoid the need to summarize outputs.

Recent work in ABC

Recent work on ABC has sought to move away from the use of summaries

- Bernton *et al.* 2017 look at Bayes like procedures based on the Wasserstein distance (get different pseudo-true value)
- Park *et al.* 2015 look at using kernel mean embeddings of distributions to also avoid the need to summarize outputs.

Several papers (Frazier *et al.* 2017, Ridgeway 2017, ...) have studied asymptotic properties of ABC

- Consider version of ABC where we accept or reject according to

$$\rho(\eta(y), \eta(y'))$$

  where $y' \sim F_\theta(\cdot)$

Then if b_0 is limit of $\eta(y)$ and $b(\theta)$ the limit of $\eta(y')$, then they've studied convergence to

$$\theta^* = \arg\inf_\theta \rho(b_0, b(\theta))$$

as $\epsilon \to 0$.

This focus is again on prediction not inference.

Discussion

What properties do we want our inference scheme to possess?

Discussion

What properties do we want our inference scheme to possess?

- Is coherence the best we can hope for or is there a form of robustness that is achievable and useful for slightly mis-specified models?

Discussion

What properties do we want our inference scheme to possess?

- Is coherence the best we can hope for or is there a form of robustness that is achievable and useful for slightly mis-specified models?
- If $G \notin \mathcal{F}$ can we ever hope to learn precisely about θ?
  If not we shouldn't use methods that converge/concentrate asymptotically.

Discussion

What properties do we want our inference scheme to possess?

- Is coherence the best we can hope for or is there a form of robustness that is achievable and useful for slightly mis-specified models?
- If $G \notin \mathcal{F}$ can we ever hope to learn precisely about θ?
  If not we shouldn't use methods that converge/concentrate asymptotically.
- Bayes linear type specification of discrepancies look attractive in most cases. Use methods that allow for this type of simple specification?

Discussion

What properties do we want our inference scheme to possess?

- Is coherence the best we can hope for or is there a form of robustness that is achievable and useful for slightly mis-specified models?
- If $G \notin \mathcal{F}$ can we ever hope to learn precisely about θ?
  If not we shouldn't use methods that converge/concentrate asymptotically.
- Bayes linear type specification of discrepancies look attractive in most cases. Use methods that allow for this type of simple specification?

Thank you for listening!