Is it possible to be a subjective Bayesian when working on 21st Century problems?

Danny Williamson[1] & Michael Goldstein[2]

[1]Statistical Science
College of Engineering, Mathematics and Physical Sciences
University of Exeter

[2]Department of Mathematical Sciences
Durham University

August 28, 2015

Motivation, Meaning and Clarity

Meaning and clarity in statistical inference

- When working as a statistician, it is not uncommon to be asked to work on the following types of problem:
- What is **the** probability that *something will happen*?
  - **Example:** What is the probability that the Earth will warm by more than 2^o by 2050?
- Quantify **the** uncertainty in *some real world quantity*
  - **Example:** Quantify the uncertainty in change in UK rainfall by 2080 due to climate change
- Are these questions meaningful?
- Is it clear to scientist and statistician what our analysis means?

Subjective Bayes

- Subjective Bayes is a route towards modest meaning and clarity for statistical analysis.
- You are uncertain about the world
- You specify your uncertainty in the form of a prior probability distribution
- You specify a probability model describing the data generating process you will use to learn about the world.
- You update your prior to your posterior uncertainty through Bayes theorem having seen the data.
- The resulting uncertainty is **yours**.
- Calculated probabilities are **your** probabilities.

21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:

21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:
  - Big data has arrived

21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:
  - Big data has arrived
  - Powerful computers and supercomputer access for analysis

21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:
  - Big data has arrived
  - Powerful computers and supercomputer access for analysis
  - Challenging science

21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:
  - Big data has arrived
  - Powerful computers and supercomputer access for analysis
  - Challenging science
  - Algorithms like MCMC/INLA allow us to do it (and the algorithm arms race continues).

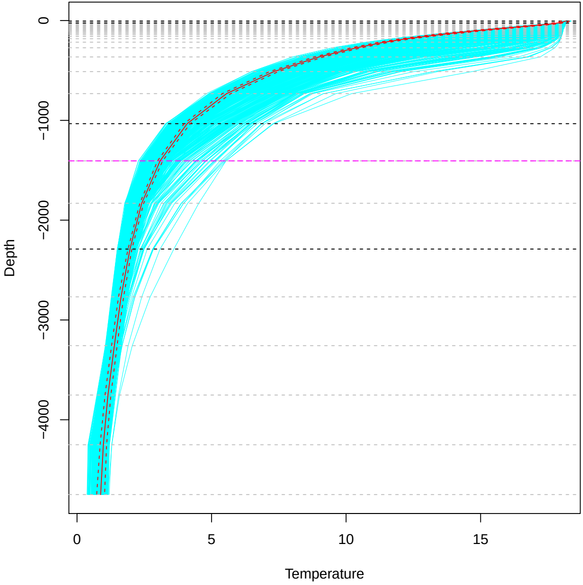
21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:
  - Big data has arrived
  - Powerful computers and supercomputer access for analysis
  - Challenging science
  - Algorithms like MCMC/INLA allow us to do it (and the algorithm arms race continues).
  - Publishing/Career pressure?

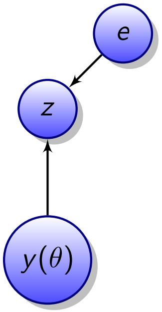
21st century Bayesian statistics

- We are all under pressure to produce novel, complex Bayesian analyses:
  - Big data has arrived
  - Powerful computers and supercomputer access for analysis
  - Challenging science
  - Algorithms like MCMC/INLA allow us to do it (and the algorithm arms race continues).
  - Publishing/Career pressure?
  - Pressure from other scientists

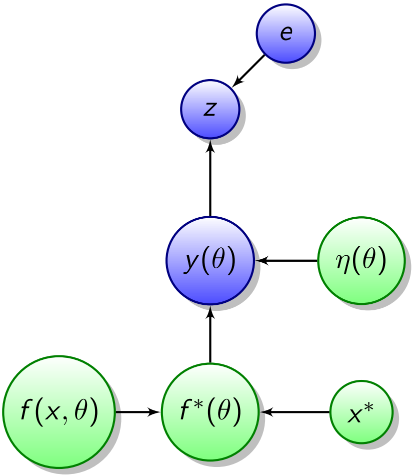
The NEMO ocean model



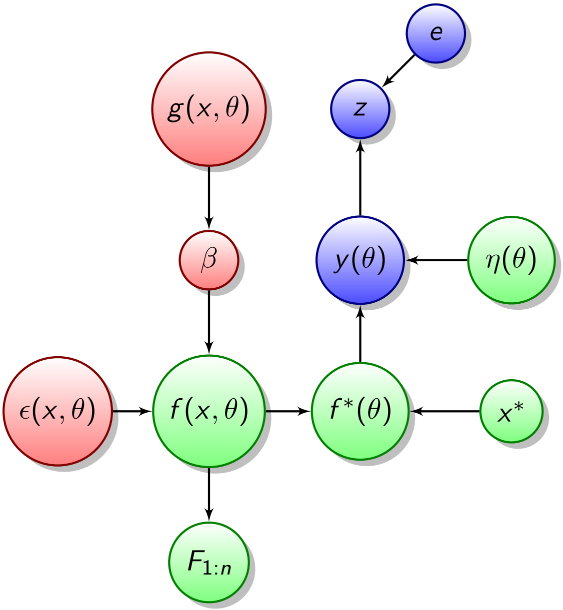
Example: Ocean model calibration



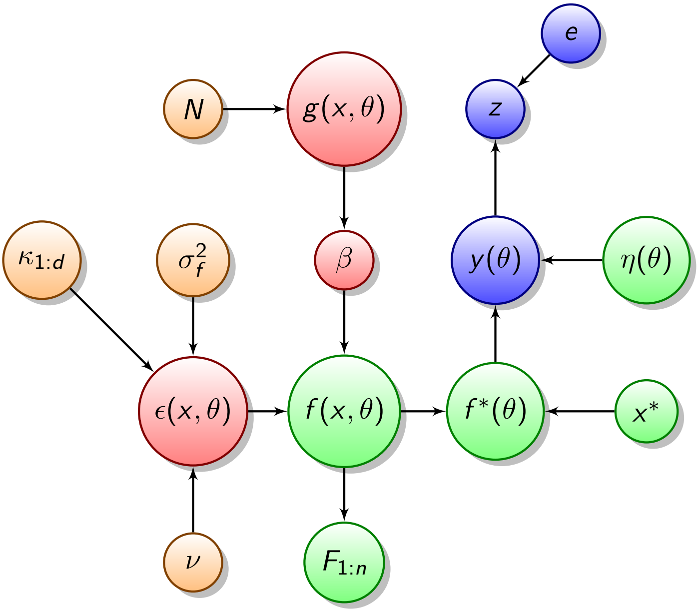
Example: Ocean model calibration



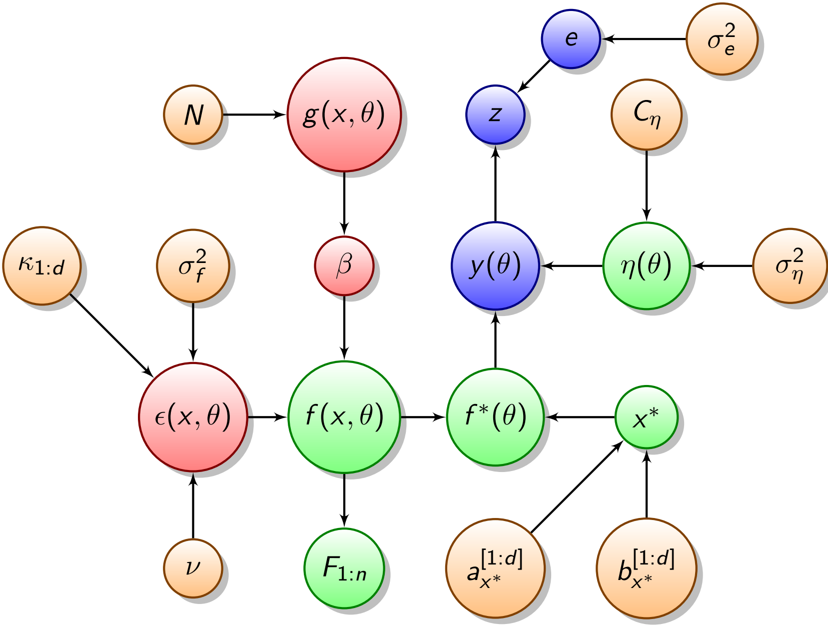
Example: Ocean model calibration



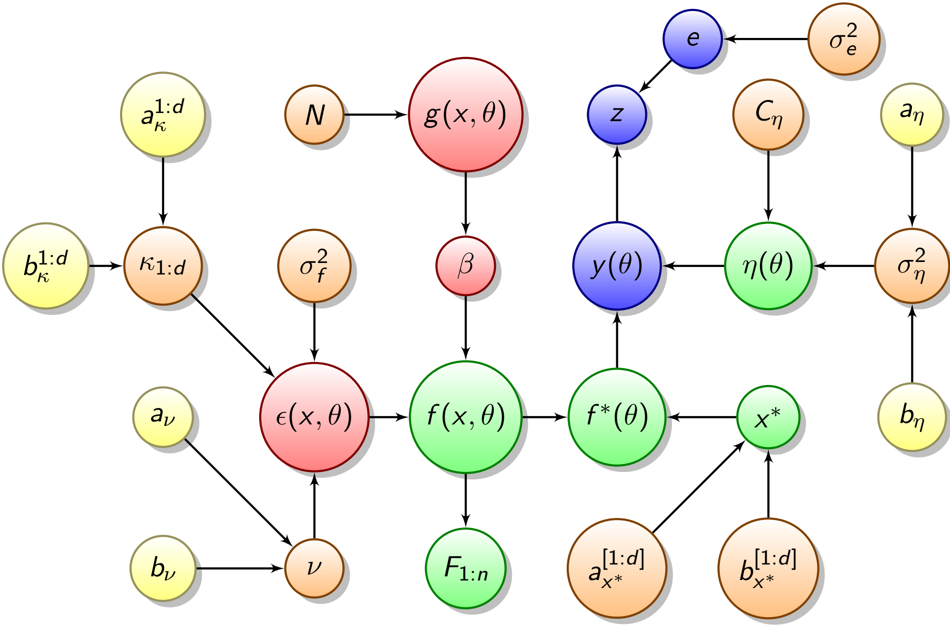
Example: Ocean model calibration



Example: Ocean model calibration



Example: Ocean model calibration



Who is the expert?

- The statistician often (all the time) plays the role of expert. Model selection, prior forms, likelihood, common parameterisations, "non-informative" priors...

Who is the expert?

- The statistician often (all the time) plays the role of expert. Model selection, prior forms, likelihood, common parameterisations, "non-informative" priors...
- We guide the scientist(s), they guide us, and sometimes pragmatism takes over.

Who is the expert?

- The statistician often (all the time) plays the role of expert. Model selection, prior forms, likelihood, common parameterisations, "non-informative" priors...
- We guide the scientist(s), they guide us, and sometimes pragmatism takes over.
- Who owns the final judgments? Both of us? Either?

Who is the expert?

- The statistician often (all the time) plays the role of expert. Model selection, prior forms, likelihood, common parameterisations, "non-informative" priors...
- We guide the scientist(s), they guide us, and sometimes pragmatism takes over.
- Who owns the final judgments? Both of us? Either?
- If not either, what meaning do we give the results?
- Is there anything we can conclude about our own uncertainty from a Bayesian analysis?

Posterior belief assessment

UNIVERSITY OF EXETER

Expectation as primitive

Definition

Your expectation, $\mathrm{E}[X]$ *for random quantity* X *is your preferred choice for the value of* c *when confronted with the random penalty* $(X - c)^2$*.*

- For any value V, you prefer $(X - \mathrm{E}[X])^2$ to $(X - V)^2$.
- Equivalently

$$\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \leq \mathrm{E}\left[(X - V)^2\right]$$

Bayes as a model for inference

- Consider the conditional expectation $\mathrm{E}[X|D]$, and the actual expectation $P_t(X)$ that we may specify at time t when we do observe D_i.

Bayes as a model for inference

- Consider the conditional expectation $\mathrm{E}[X|D]$, and the actual expectation $P_t(X)$ that we may specify at time t when we do observe D_i.
- These are not the same!

Bayes as a model for inference

- Consider the conditional expectation $\mathrm{E}[X|D]$, and the actual expectation $P_t(X)$ that we may specify at time t when we do observe D_i.
- These are not the same!
- $\mathrm{E}[X|D]$ is a bet that will be called off unless the conditioning event occurs (one you make now).

Bayes as a model for inference

- Consider the conditional expectation $\mathrm{E}[X|D]$, and the actual expectation $P_t(X)$ that we may specify at time t when we do observe D_i.
- These are not the same!
- $\mathrm{E}[X|D]$ is a bet that will be called off unless the conditioning event occurs (one you make now).
- $P_t(X)$ is the posterior probability that you are free to assign at time t having seen the event.

Bayes as a model for inference

- Consider the conditional expectation $\mathrm{E}[X|D]$, and the actual expectation $P_t(X)$ that we may specify at time t when we do observe D_i.
- These are not the same!
- $\mathrm{E}[X|D]$ is a bet that will be called off unless the conditioning event occurs (one you make now).
- $P_t(X)$ is the posterior probability that you are free to assign at time t having seen the event.
- There is no obvious formal relationship between $\mathrm{E}[X|D]$ and $P_t(X)$ whatsoever!
- As such, conditional probability can **only** be interpreted as a *model* for real world inference.

Temporal Sure Preference

- To relate the model (conditional probability) to reality (actual posterior prevision) requires something more.
- Suppose that we have two random penalties A and B that we must choose between

Definition

We say that we have a **sure preference** *for A over B at future time t if we are sure now that, at time t, we will prefer A to B.*

The Temporal Sure Preference Principle

If we have a sure preference for A over B at t, then we should not have a preference for B over A now

Interpreting our modelling judgements

- Let our quantities of interest be vector y with relevant data z.
- Define our current modelling judgements as the set J_0
  - Form of prior and likelihood
  - Hyper-parameters
  - Computational methods

Interpreting our modelling judgements

- Let our quantities of interest be vector y with relevant data z.
- Define our current modelling judgements as the set J_0
  - Form of prior and likelihood
  - Hyper-parameters
  - Computational methods
- **We do not need to believe all of the probability statements implied by $\mathbf{J_0}$**

Interpreting our modelling judgements

- Let our quantities of interest be vector y with relevant data z.
- Define our current modelling judgements as the set J_0
  - Form of prior and likelihood
  - Hyper-parameters
  - Computational methods
- **We do not need to believe all of the probability statements implied by $\mathbf{J_0}$**
- We view J_0 as "representative enough" of the structure of the problem and any scientific judgements that we do hold so that we view $\mathrm{E}[y|z; J_0]$ as *informative* for our posterior prevision for y.
- J_0 will not be unique.

Alternative judgements

- Changes to J_0 may represent our scientific insights or beliefs in different ways.

Alternative judgements

- Changes to J_0 may represent our scientific insights or beliefs in different ways.
- Or in ways that are subtle so that we would still view the analysis as informative for y.

Alternative judgements

- Changes to J_0 may represent our scientific insights or beliefs in different ways.
- Or in ways that are subtle so that we would still view the analysis as informative for y.
- For example, certain prior choices made for convenience (e.g. conjugacy) may be reconsidered here.

Alternative judgements

- Changes to J_0 may represent our scientific insights or beliefs in different ways.
- Or in ways that are subtle so that we would still view the analysis as informative for y.
- For example, certain prior choices made for convenience (e.g. conjugacy) may be reconsidered here.
- Consider a finite (for now) set of alternative judgements $J_1, \ldots, J_k$ that we view as potentially informative for our prevision for y if we were to conduct the Bayesian analysis under each and compute $\mathrm{E}[y|z; J_i]$ for $i = 1, \ldots, k$.
- What does the existence of alternative judgements imply for our uncertainty?

Interpretation

- There is no need to posit a true but unknown judgement set J^*.
- Could any expert have infinite resolution and own all of J^*?
- If so, and we just don't have enough time to elicit it, is this useful?

Interpretation

- There is no need to posit a true but unknown judgement set J^*.
- Could any expert have infinite resolution and own all of J^*?
- If so, and we just don't have enough time to elicit it, is this useful?
- Our view is that an expert can hold previsions for a small set of key quantities (y) and that a Bayesian analysis under any $J_0, J_1, \ldots, J_k$ is informative, in some way, for that prevision.
- This view uses probability in two ways:
  1. What we actually believe: our previsions.
  2. A rich and powerful modelling language for transforming judgements into previsions.

Posterior belief assessment

- TSP implies

$$\mathrm{E}\left[(y_i - P_t(y_i))^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}\left[y_i | z; J_0\right])^2\right]$$

Posterior belief assessment

- TSP implies

$$\mathrm{E}\left[(y_i - P_t(y_i))^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_0\right])^2\right]$$

- That we have chosen J_0 over any $J_1, \ldots, J_k$, **may** lead us to the view that

$$\mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_0\right])^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_i\right])^2\right] \qquad i = 1, \ldots, k.$$

Posterior belief assessment

- TSP implies

$$\mathrm{E}\left[(y_i - P_t(y_i))^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_0\right])^2\right]$$

- That we have chosen J_0 over any $J_1, \ldots, J_k$, **may** lead us to the view that

$$\mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_0\right])^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_i\right])^2\right] \qquad i = 1, \ldots, k.$$

- However, is there some vector valued function $\Gamma(z)$ so that

$$\mathrm{E}\left[(y_i - \Gamma_i(z))^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}\left[y_i|z; J_0\right])^2\right]?$$

Posterior belief assessment

Define $\mathcal{G}$ to be the vector $(\mathrm{E}[y|z;J_0], \mathrm{E}[y|z;J_1], \ldots, \mathrm{E}[y|z;J_k]) = (\mathcal{G}_1, \ldots, \mathcal{G}_{k+1})$. Let $\mathcal{G}_0$ be the unit constant.

Posterior belief assessment

Define $\mathcal{G}$ to be the vector $(\mathrm{E}[y|z;J_0], \mathrm{E}[y|z;J_1], \ldots, \mathrm{E}[y|z;J_k]) = (\mathcal{G}_1, \ldots, \mathcal{G}_{k+1})$. Let $\mathcal{G}_0$ be the unit constant.

Theorem

Let

$$\mathrm{E}_{\mathcal{G}}[y] = \mathrm{E}[y] + \mathrm{Cov}[y, \mathcal{G}] \, \mathrm{Var}[\mathcal{G}]^{-1} (\mathcal{G} - \mathrm{E}[\mathcal{G}]). \tag{1}$$

Then

(i) $\mathrm{E}_{\mathcal{G}}[y]$ *is at least as close to* y *as* $\mathrm{E}[y|z;J_0]$. *Equivalently, for each* i,

$$\mathrm{E}\left[(y_i - \mathrm{E}_{\mathcal{G}}[y_i])^2\right] \leq \mathrm{E}\left[(y_i - \mathrm{E}[y_i|z;J_0])^2\right].$$

where $\mathrm{E}_{\mathcal{G}}[y_i]$ *is the* i*th component of* $\mathrm{E}_{\mathcal{G}}[y]$.

(ii) $\mathrm{E}_{\mathcal{G}}[y]$ *is at least as close to* $P_t(y)$ *as* $\mathrm{E}[y|z;J_0]$. *Equivalently, for each* i,

$$\mathrm{E}\left[(P_t(y_i) - \mathrm{E}_{\mathcal{G}}[y_i])^2\right] \leq \mathrm{E}\left[(P_t(y_i) - \mathrm{E}[y_i|z;J_0])^2\right].$$

An algorithm for computing Posterior belief assessments

- Assume we can elicit $\mathrm{E}[y], \mathrm{Var}[y]$ and $\mathrm{E}[z|\hat{y}], \mathrm{Var}[z|\hat{y}]$

Step 1 Sample a value $\hat{y}$ and $\hat{z}$ from $(\mathrm{E}[y], \mathrm{Var}[y])$ and $(\mathrm{E}[z|\hat{y}], \mathrm{Var}[z|\hat{y}])$ respectively.

Step 2 Use the full Bayes machinery to compute each $\mathrm{E}[y|\hat{z}; J_0]$, $\mathrm{E}[y|\hat{z}; J_1], \ldots$, and use them to form $\hat{\mathcal{G}}$.

Step 3 Repeat this process to obtain a large number, N, of sample pairs $(\hat{y}_1, \hat{\mathcal{G}}_1), \ldots, (\hat{y}_N, \hat{\mathcal{G}}_N)$.

Step 4 Assess $\mathrm{E}[\mathcal{G}]$, $\mathrm{Var}[\mathcal{G}]$ and $\mathrm{Cov}[y, \mathcal{G}]$ by computing the sample means and variances of the $\hat{\mathcal{G}}s$ and their covariance with the $\hat{y}$s.

UNIVERSITY OF EXETER

Coexchangeable classes and posterior belief assessment

Alternative judgements

- The collection of possible alternative judgements, $J_1, J_2, \ldots$ may (will) not be finite.

Alternative judgements

- The collection of possible alternative judgements, $J_1, J_2, \ldots$ may (will) not be finite.
- Some alternative judgements may lead to entirely different models and a great deal of extra work we don't have time for.

Alternative judgements

- The collection of possible alternative judgements, $J_1, J_2, \ldots$ may (will) not be finite.
- Some alternative judgements may lead to entirely different models and a great deal of extra work we don't have time for.
- Running a large number of alternative Bayesian analyses (say by MCMC) may be computationally challenging or infeasible.

Alternative judgements

- The collection of possible alternative judgements, $J_1, J_2, \dots$ may (will) not be finite.
- Some alternative judgements may lead to entirely different models and a great deal of extra work we don't have time for.
- Running a large number of alternative Bayesian analyses (say by MCMC) may be computationally challenging or infeasible.
- We collect judgements $J_1, J_2, \dots$ into classes $C_1, \dots, C_k$ so that we judge each collection $\mathrm{E}\left[y|z; J_{[C_l]1}\right], \mathrm{E}\left[y|z; J_{[C_l]2}\right], \mathrm{E}\left[y|z; J_{[C_l]3}\right], \dots$ to be infinite **second order exchangeable**.

Co-exchangeable classes of exchangeable judgements

$X_1, X_2, X_3 \ldots$ are second order exchangeable if

$$\mathrm{E}[X_i] = \mu; \qquad \mathrm{Var}[X_i] = \Sigma \quad \forall i; \qquad \mathrm{Cov}[X_i, X_j] = \Gamma \quad \forall i \neq j.$$

Co-exchangeable classes of exchangeable judgements

$X_1, X_2, X_3 \ldots$ are second order exchangeable if

$$\mathrm{E}[X_i] = \mu; \qquad \mathrm{Var}[X_i] = \Sigma \quad \forall i; \qquad \mathrm{Cov}[X_i, X_j] = \Gamma \quad \forall i \neq j.$$

Also, for an infinite second order exchangeable collection we have

$$X_j = \mathcal{M}(X) + \mathcal{R}_j(X)$$

with

$$\mathrm{E}[\mathcal{M}(X)] = \mu; \qquad \mathrm{Var}[\mathcal{M}(X)] = \Gamma;$$
$$\mathrm{E}[R_j(X)] = 0; \qquad \mathrm{Var}[R_j(X)] = \Sigma - \Gamma;$$
$$\mathrm{Cov}[\mathcal{M}(X), \mathcal{R}_j(X)] = \mathrm{Cov}[R_j(X), R_k(X)] = 0 \quad \forall j, k \neq j$$

Co-exchangeable classes of exchangeable judgements

- Further, if we are willing to impose exchangeability on each class, it is quite natural to impose co-exchangeability between the classes.
- Co-exchangeability implies

$$\mathrm{Cov}\left[\mathrm{E}\left[y|z;J_{[C_k]i}\right],\mathrm{E}\left[y|z;J_{[C_l]j}\right]\right]=\Sigma_{kl}\quad\forall i,j.$$

- We now have a rich structure from which we can make inferences.

Co-exchangeable classes of exchangeable judgements

- Further, if we are willing to impose exchangeability on each class, it is quite natural to impose co-exchangeability between the classes.
- Co-exchangeability implies

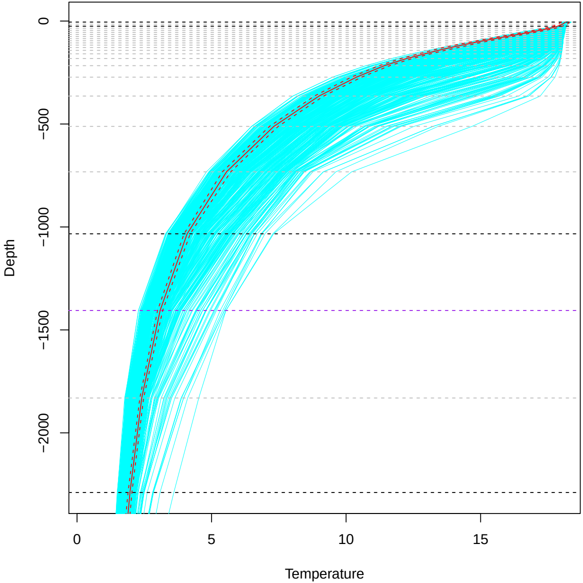
$$\mathrm{Cov}\left[\mathrm{E}\left[y|z;J_{[C_k]i}\right],\mathrm{E}\left[y|z;J_{[C_l]j}\right]\right]=\Sigma_{kl}\quad\forall i,j.$$

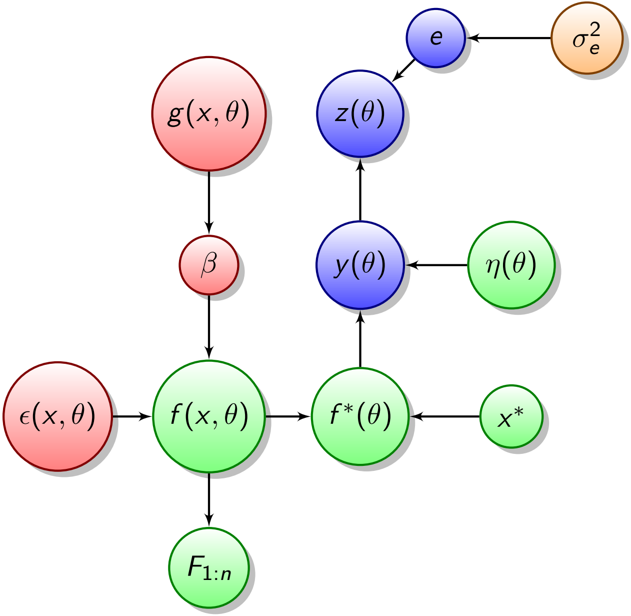
- We now have a rich structure from which we can make inferences.
- In particular, if we can list all of the classes, $C_1,\ldots,C_k$, we can design a handful of extra Bayesian analyses to run under judgements $J_{[C_{l(1)}]j(1)},\ldots,J_{[C_{l(n)}]j(n)}$.
- We can then use $\mathrm{E}\left[y|z;J_{[C_{l(1)}]j(1)}\right],\ldots,\mathrm{E}\left[y|z;J_{[C_{l(n)}]j(n)}\right]$ to make inference about the conditional expectations output from any alternative Bayesian analysis under different judgements.

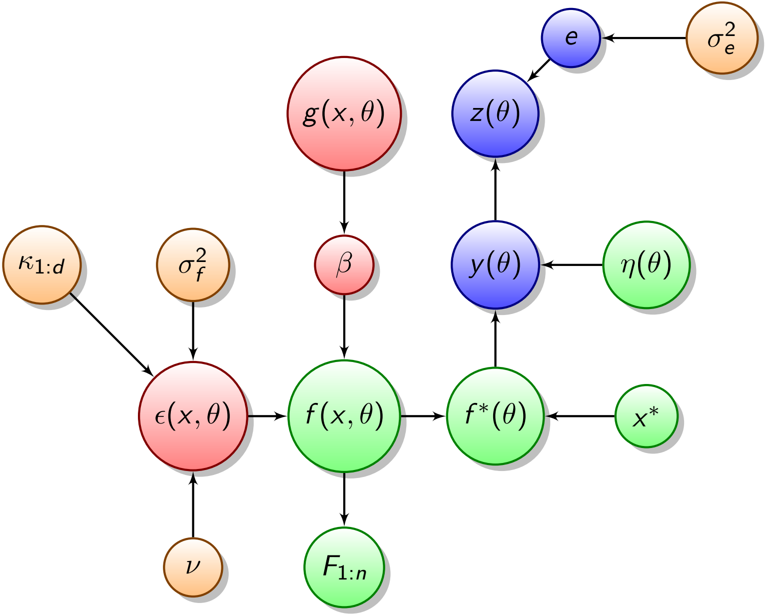
UNIVERSITY OF EXETER

Application: Calibrating NEMO

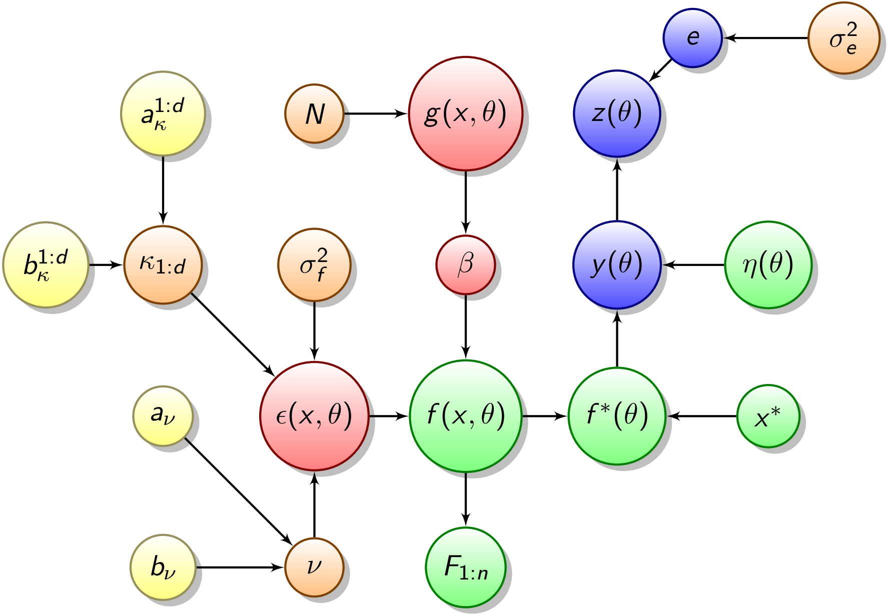
The NEMO ocean model

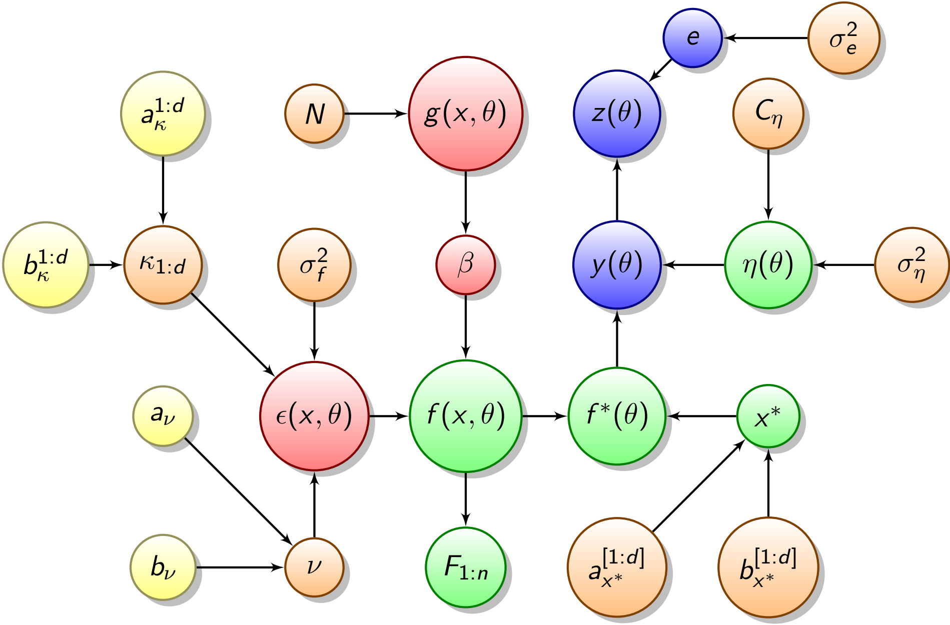



e
σ_e^2
$g(x, \theta)$
$z(\theta)$
β
$y(\theta)$
$\eta(\theta)$
$\epsilon(x, \theta)$
$f(x, \theta)$
$f^*(\theta)$
x^*
$F_{1:n}$

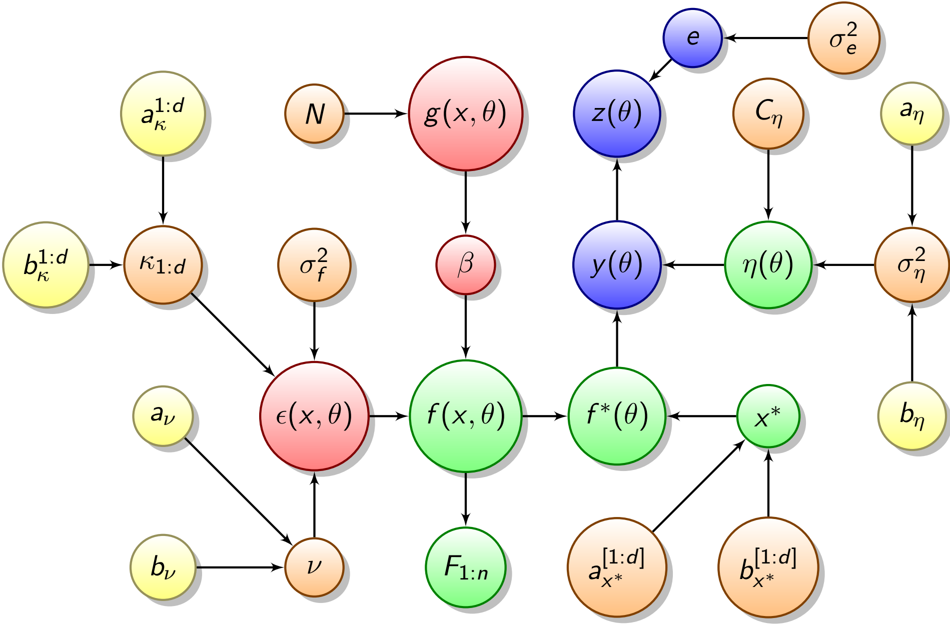


σ_e^2
e
$g(x,\theta)$
$z(\theta)$
$\kappa_{1:d}$
σ_f^2
β
$y(\theta)$
$\eta(\theta)$
$\epsilon(x,\theta)$
$f(x,\theta)$
$f^*(\theta)$
x^*
ν
$F_{1:n}$


σ_e^2
e
$a_\kappa^{1:d}$
N
$g(x,\theta)$
$z(\theta)$
$b_\kappa^{1:d}$
$\kappa_{1:d}$
σ_f^2
β
$y(\theta)$
$\eta(\theta)$
a_ν
$\epsilon(x,\theta)$
$f(x,\theta)$
$f^*(\theta)$
x^*
b_ν
ν
$F_{1:n}$



e
σ_e^2
$a_\kappa^{1:d}$
N
$g(x,\theta)$
$z(\theta)$
C_η
$b_\kappa^{1:d}$
$\kappa_{1:d}$
σ_f^2
β
$y(\theta)$
$\eta(\theta)$
σ_η^2
a_ν
$\epsilon(x,\theta)$
$f(x,\theta)$
$f^*(\theta)$
x^*
b_ν
ν
$F_{1:n}$
$a_{x^*}^{[1:d]}$
$b_{x^*}^{[1:d]}$



e
σ_e^2
$a_\kappa^{1:d}$
N
$g(x,\theta)$
$z(\theta)$
C_η
a_η
$b_\kappa^{1:d}$
$\kappa_{1:d}$
σ_f^2
β
$y(\theta)$
$\eta(\theta)$
σ_η^2
a_ν
$\epsilon(x,\theta)$
$f(x,\theta)$
$f^*(\theta)$
x^*
b_η
b_ν
ν
$F_{1:n}$
$a_{x^*}^{[1:d]}$
$b_{x^*}^{[1:d]}$

Performing different analyses

- We defined 360 combinations of alternative judgements in 6 co-exchangeable classes.

Performing different analyses

- We defined 360 combinations of alternative judgements in 6 co-exchangeable classes.
- MCMC takes about 1 hour per run for this size model and data.

Performing different analyses

- We defined 360 combinations of alternative judgements in 6 co-exchangeable classes.
- MCMC takes about 1 hour per run for this size model and data.
- We run 73 different experiments (including J_0) across the classes to get $\mathrm{E}\left[y|z; J_{[C_j]k}\right]$, taking just over 18 hours on a 4 core machine.

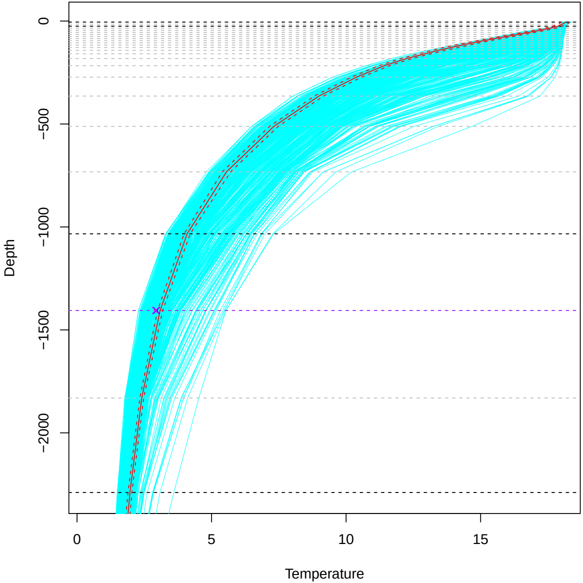
Performing different analyses

- We defined 360 combinations of alternative judgements in 6 co-exchangeable classes.
- MCMC takes about 1 hour per run for this size model and data.
- We run 73 different experiments (including J_0) across the classes to get $\mathrm{E}\left[y|z; J_{[C_j]k}\right]$, taking just over 18 hours on a 4 core machine.
- We used our algorithm to get all of the required quantities to do the PBA

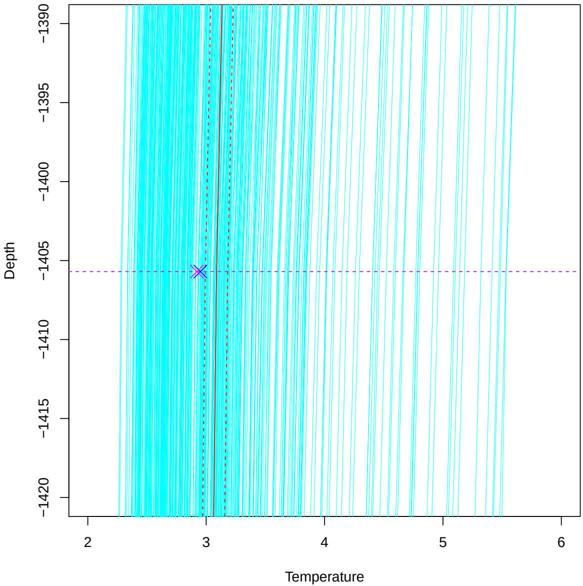
Performing different analyses

- We defined 360 combinations of alternative judgements in 6 co-exchangeable classes.
- MCMC takes about 1 hour per run for this size model and data.
- We run 73 different experiments (including J_0) across the classes to get $\mathrm{E}\left[y|z; J_{[C_j]k}\right]$, taking just over 18 hours on a 4 core machine.
- We used our algorithm to get all of the required quantities to do the PBA
- $\mathrm{E}\left[y|z; J_0\right] = 2.951^o$C. $\mathrm{E}_{\mathcal{G}}\left[y\right] = 2.921^o$C.
- The uncertainty reduction achieved through completing here was at least 14%.

The NEMO ocean model



The NEMO ocean model



Conclusions

- If the problem is important enough, the answer to the question "whose judgements are these?" is crucial.

Conclusions

- If the problem is important enough, the answer to the question "whose judgements are these?" is crucial.
- Our Bayesian machinery can only produce our judgements if J_0 is the only possible set of judgements we could hold and really represents what we think (how the posterior is actually computed is in J_0 here).

Conclusions

- If the problem is important enough, the answer to the question "whose judgements are these?" is crucial.
- Our Bayesian machinery can only produce our judgements if J_0 is the only possible set of judgements we could hold and really represents what we think (how the posterior is actually computed is in J_0 here).
- We contend that this will never be true for complex problems.

Conclusions

- If the problem is important enough, the answer to the question "whose judgements are these?" is crucial.
- Our Bayesian machinery can only produce our judgements if J_0 is the only possible set of judgements we could hold and really represents what we think (how the posterior is actually computed is in J_0 here).
- We contend that this will never be true for complex problems.
- We argued that we are only interested in a handful of posterior expectations and that alternative representations of our judgements exist to help us get to these.

Conclusions

- We showed that in the presence of alternative judgements, the one off posterior expectation under $\mathrm{E}[y|z; J_0]$ does not represent your beliefs.

Conclusions

- We showed that in the presence of alternative judgements, the one off posterior expectation under $\mathrm{E}[y|z; J_0]$ does not represent your beliefs.
- We used Bayesian analyses under alternative judgements to derive posterior belief assessments, $\mathrm{E}_{\mathcal{G}}[y]$, that we proved were closer to posterior prevision than our original analysis.

Conclusions

- We showed that in the presence of alternative judgements, the one off posterior expectation under $\mathrm{E}[y|z; J_0]$ does not represent your beliefs.
- We used Bayesian analyses under alternative judgements to derive posterior belief assessments, $\mathrm{E}_{\mathcal{G}}[y]$, that we proved were closer to posterior prevision than our original analysis.
- We used exchangeability to allow us to perform a small set of further Bayesian analyses that allowed us to compute $\mathrm{E}_{\mathcal{G}}[y]$.

Williamson, D. Goldstein M. (2015),
Posterior belief assessment: extracting meaningful subjective judgements from Bayesian analyses with complex statistical models,
Bayesian Analysis, To Appear (December Issue).

UNIVERSITY OF EXETER

Bayesian model averaging (BMA)

- Put a probability distribution over an alternative set of models $M_1, \ldots, M_n$
- Then, for some real world quantity y, given data D,

$$p(y|D) = \int p(y|D, M)p(M|D)dM$$

UNIVERSITY OF EXETER

Bayesian model averaging (BMA)

- Put a probability distribution over an alternative set of models $M_1, \dots, M_n$
- Then, for some real world quantity y, given data D,

$$p(y|D) = \int p(y|D, M)p(M|D)dM$$

- BMA suffers all of the same issues as the original analysis!
  - How do we elicit $p(M)$ and what would it mean? Is there a true model?
  - If $p(M)$ can't be owned, how do we extract meaning?
  - How do we avoid the infinite regress (alternative versions of our distribution over the models)?

UNIVERSITY OF EXETER