Novel Uses of Approximate Bayesian Computation for Prior Choice

Associate Professor Chris Drovandi

School of Mathematical Sciences
Queensland University of Technology
Austraian Centre of Excellence for Mathematical and Statistical Frontiers

Joint work with David Nott, Kerrie Mengersen, Michael Evans, Wang Xueou

Motivation

- Approximate Bayesian computation (ABC) is used to sample from an approximate posterior when the likelihood is intractable.
- Simple ABC methods estimate a posterior distribution based on simulation from a joint distribution on parameters and data.
- Repeated ABC approximations of conditionals from the joint sample is easy.
- These repeated ABC approximations of conditionals can be of:
  - Parameters given data (posterior distributions)
  - Data given parameters (approximate the data model).

We show how these features make ABC useful in prior choice.

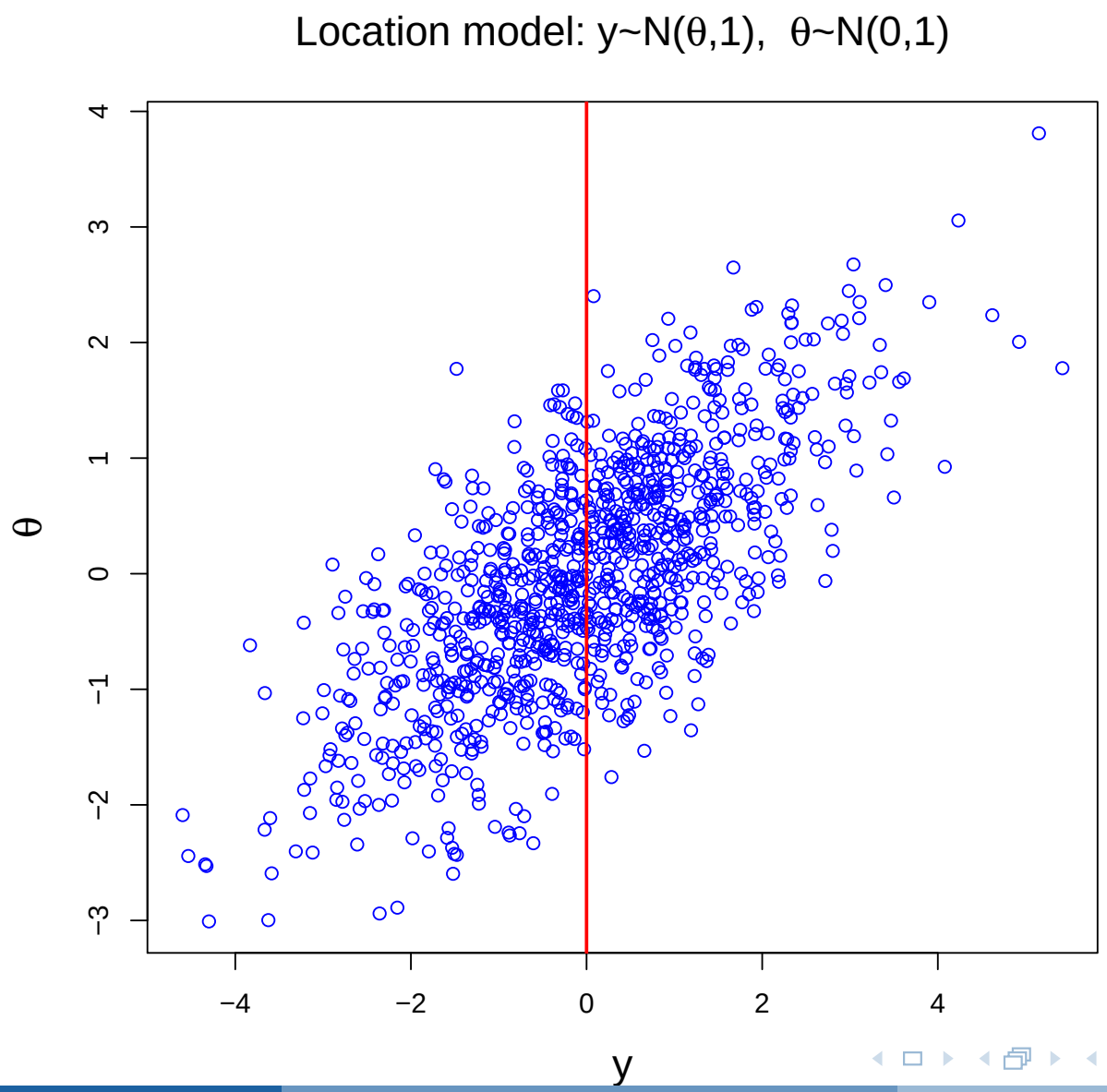
Approximate Bayesian Computation

- Data y, parameter θ, data model $p(y|\theta)$, prior $p(\theta)$.
- y_{obs} is observed and then $p(\theta|y_{obs}) \propto p(\theta)p(y_{obs}|\theta)$.
- Let $d(\cdot,\cdot)$ be a distance defined in the data space and $\epsilon > 0$ be a tolerance.

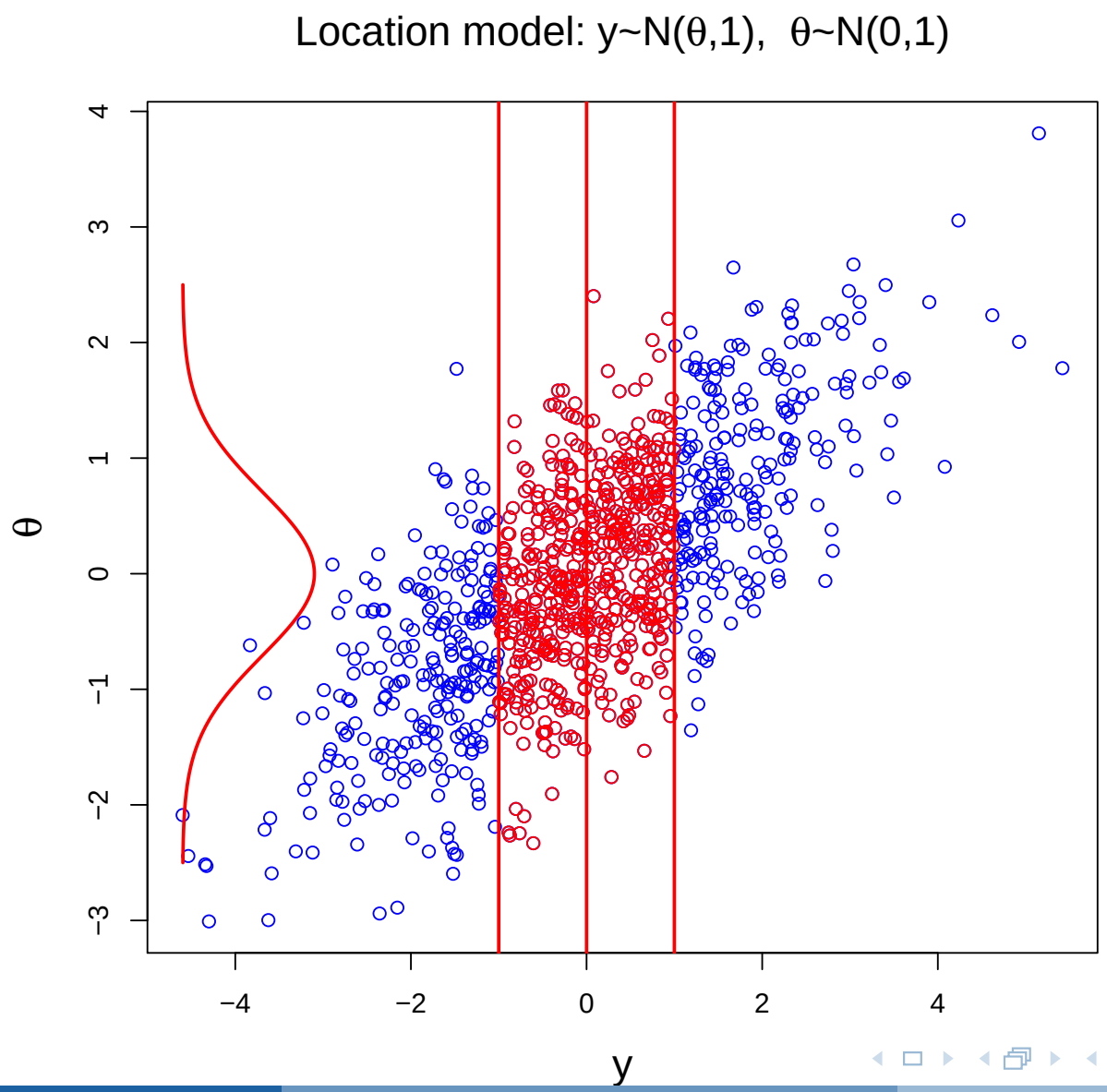
Simple ABC: condition a joint sample (θ, y) on $d(y, y_{obs}) < \epsilon$ rather than on $y = y_{obs}$ (Pritchard *et al.*, 1999)

- The distance is often constructed by reducing the data to a summary $S = S(y) \sim p(S|\theta)$ informative about θ and then $d(\cdot,\cdot)$ is defined in the space of summary statistics.

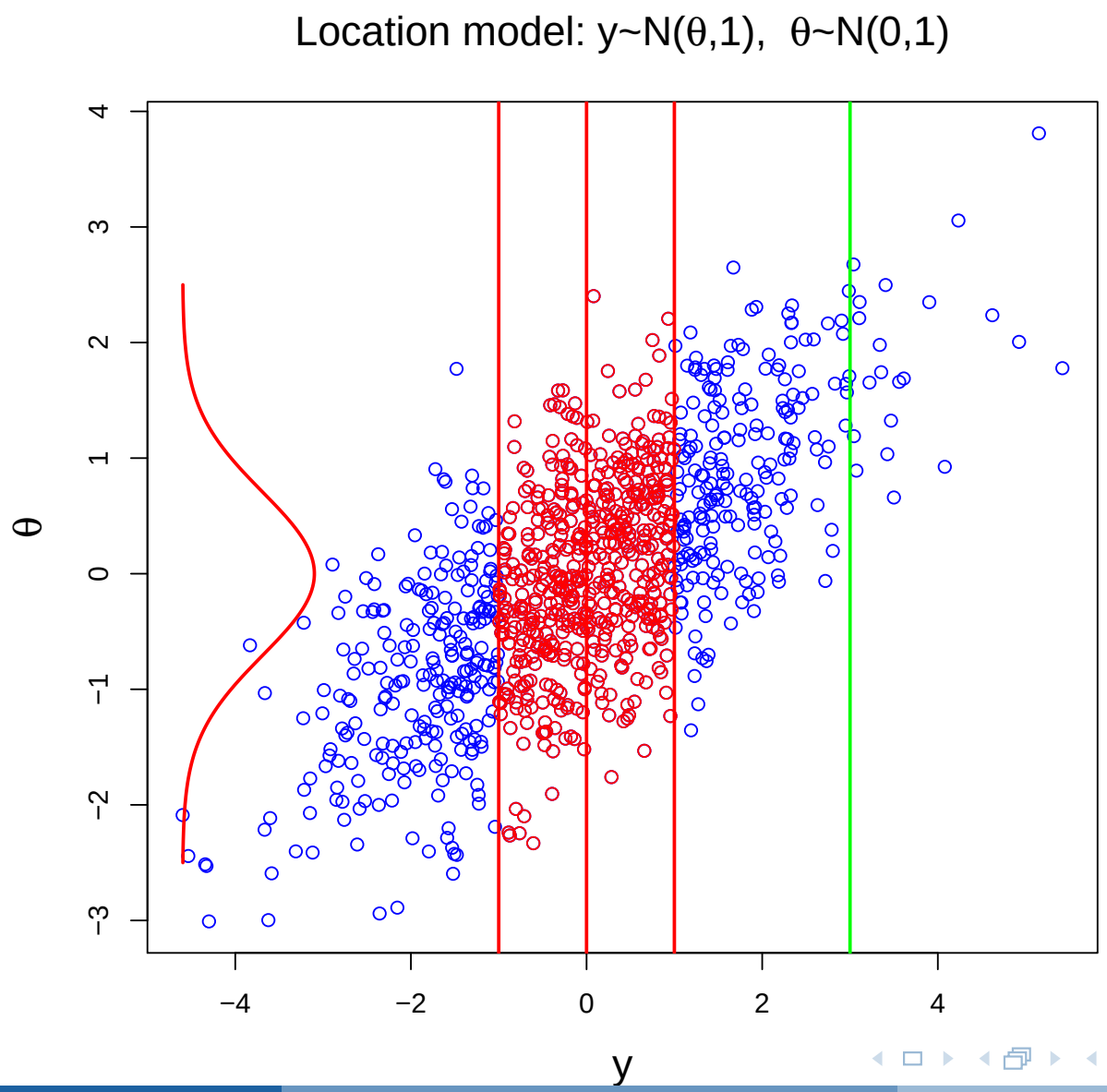
Repeated approximations of the posterior



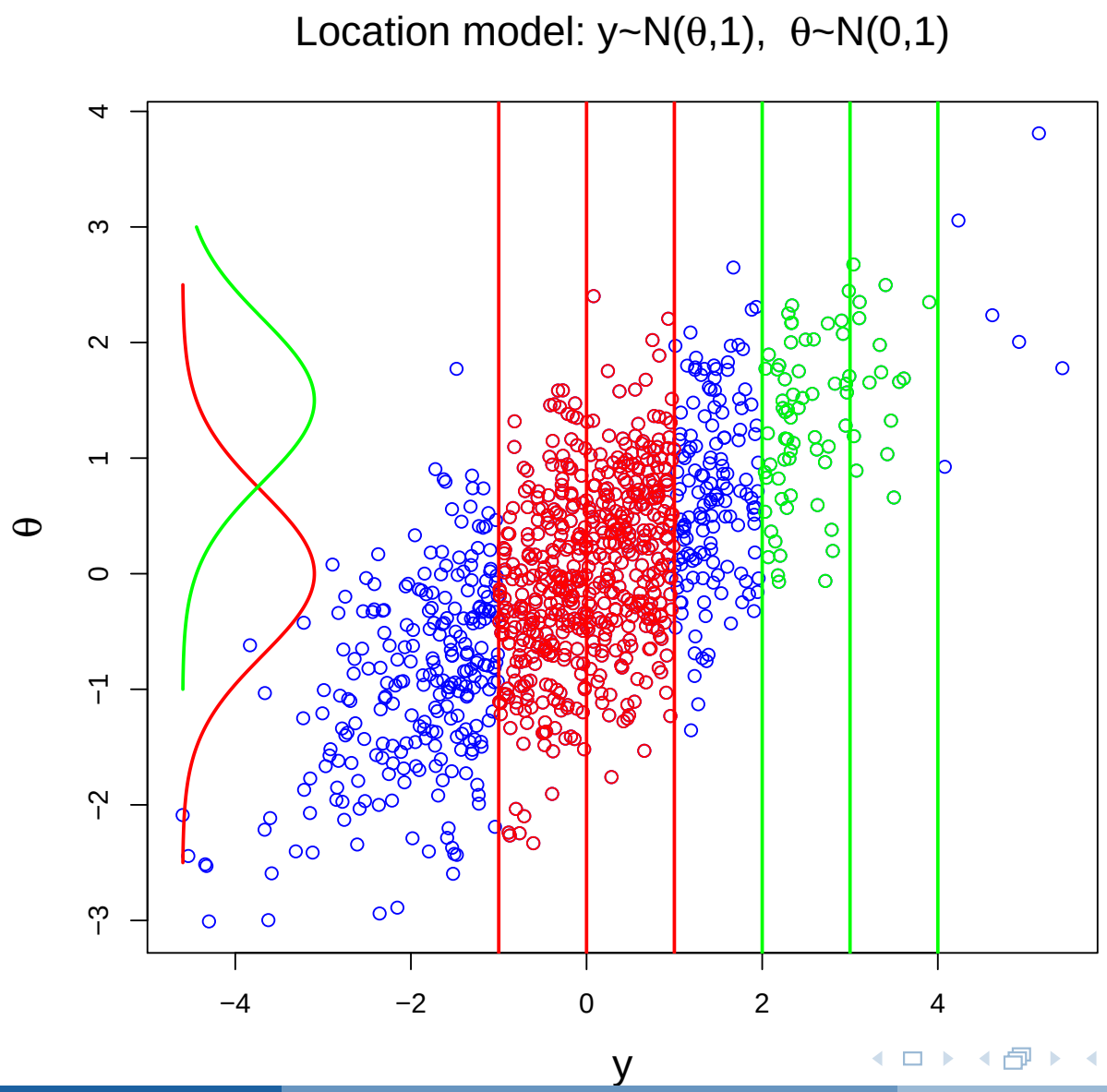
Repeated approximations of the posterior



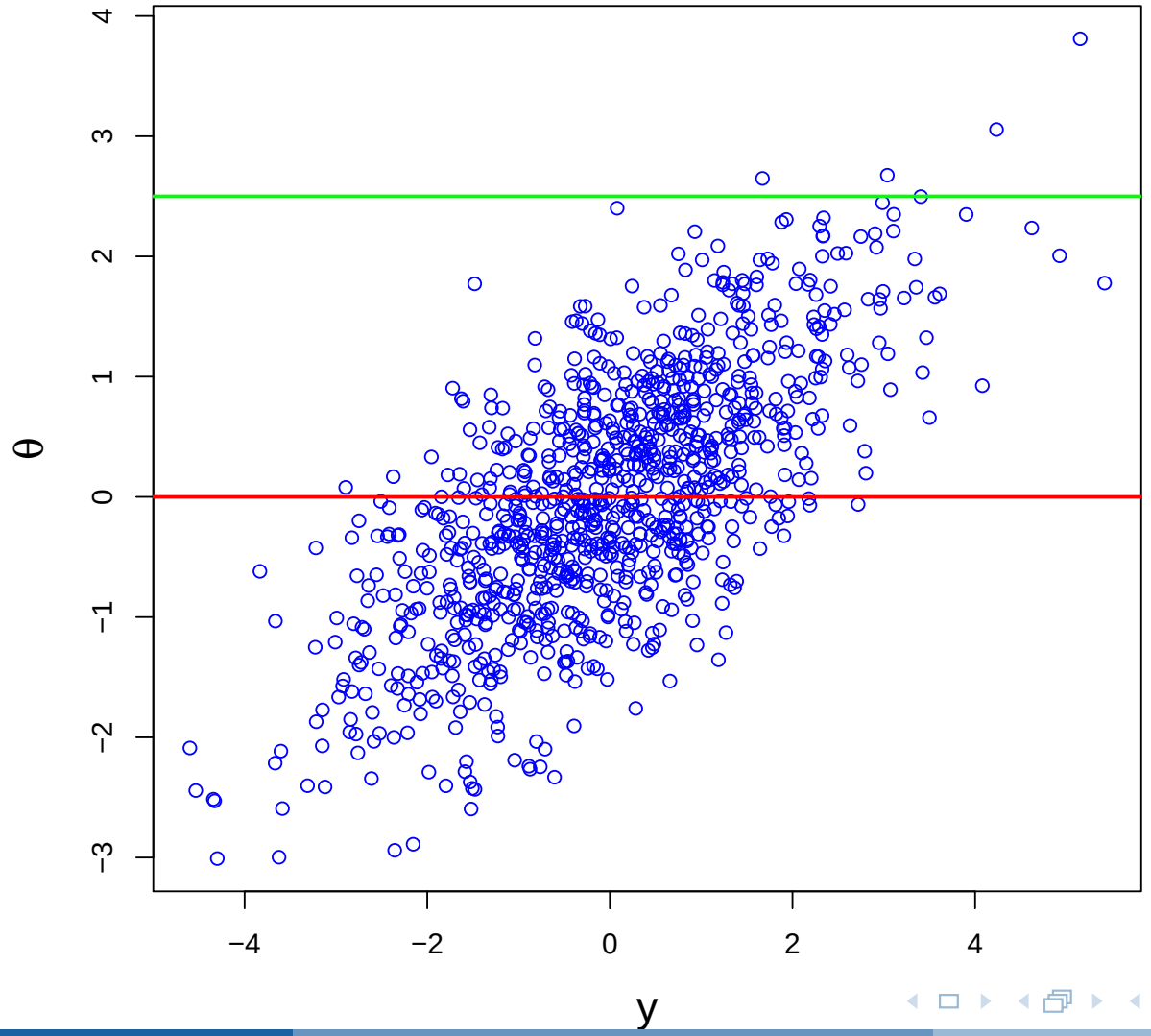
Repeated approximations of the posterior



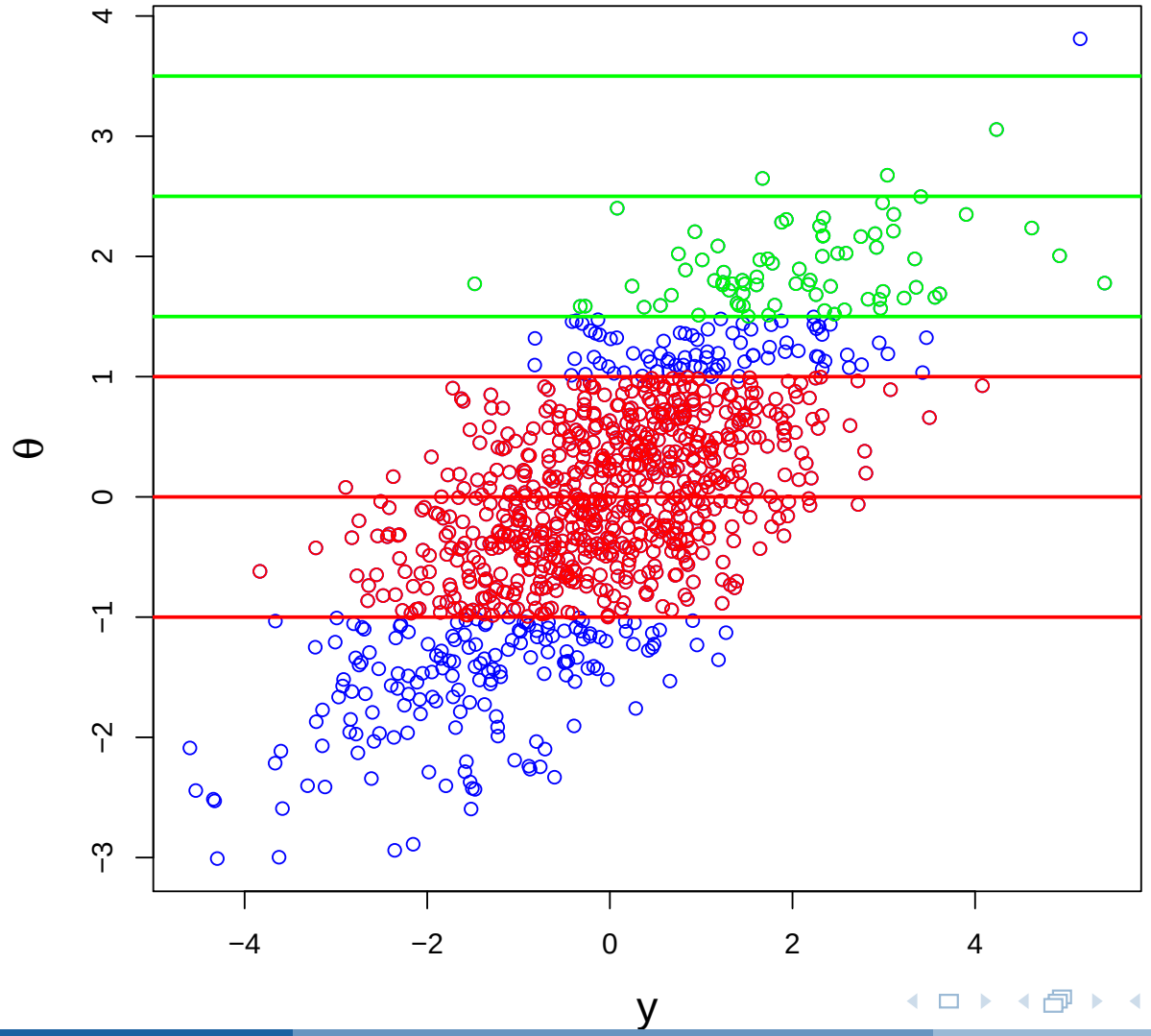
Repeated approximations of the posterior



Repeated approximations of the data model



Repeated approximations of the data model



Regression adjustment ABC

Beaumont *et al.*, 2002, Blum *et al.*, 2010

- We use *simple* ABC methods since state of the art methods lose the property of allowing repeated approximations without more simulations.
- Simple ABC algorithms can be very inefficient if the prior is diffuse.
- Regression postprocessing adjustments which allow the ABC tolerance ϵ to be large are helpful.

Regression ABC

Beaumont *et al.*, 2002, Blum *et al.*, 2010

- Suppose $(\theta_i, S_i) \sim p(\theta)p(S|\theta)$ and the regression model

$$\theta_i = \mu(S_i) + \sigma(S_i)\eta_i$$

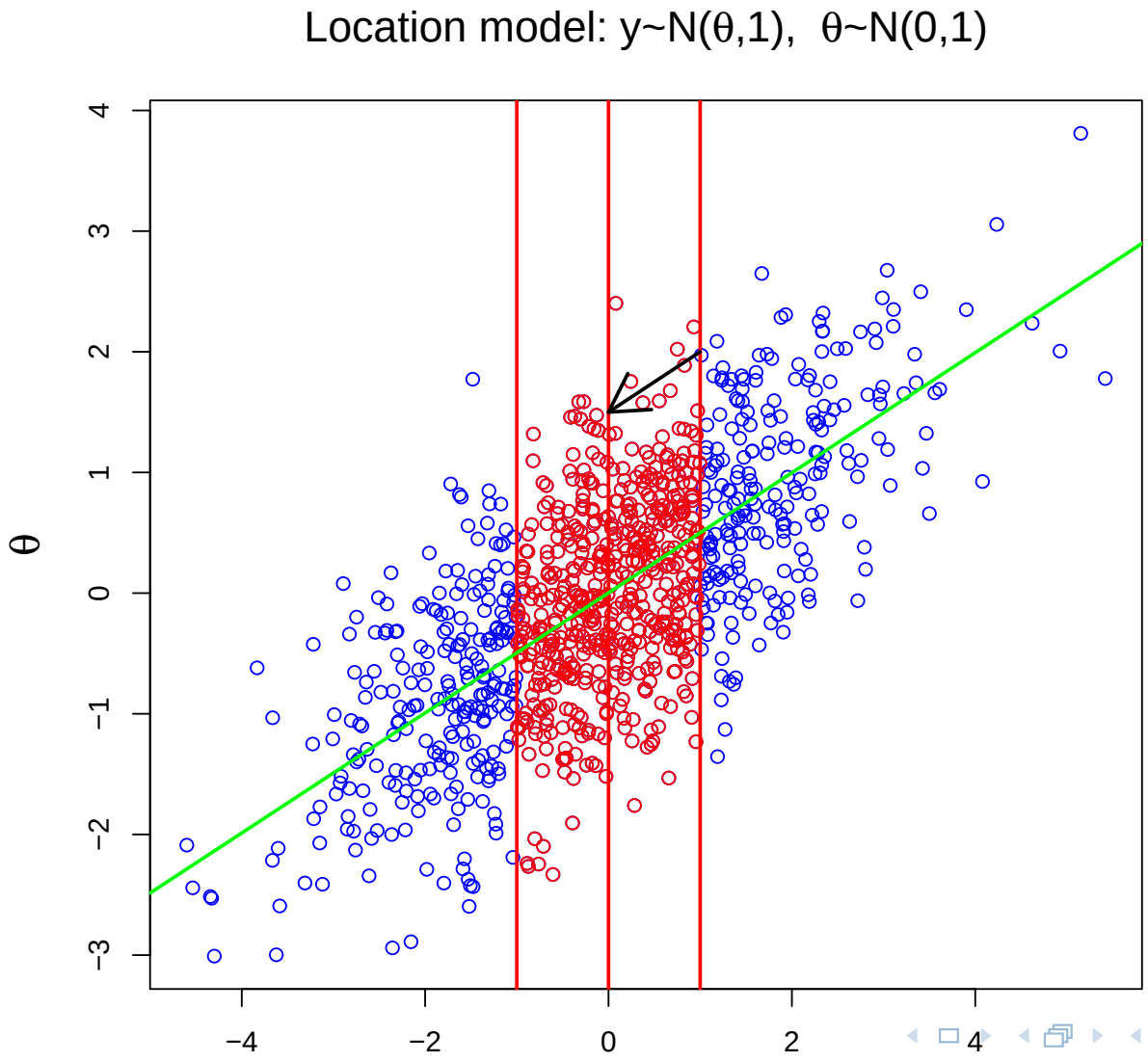
  where $\mu(\cdot)$ and $\sigma(\cdot)$ are flexible mean and standard deviation functions, η_i are independent, mean zero, variance one.
- The empirical residuals are $e_i = \hat{\sigma}(S_i)^{-1}(\theta_i - \hat{\mu}(S_i))$.
- Replace the original samples θ_i with the fitted mean at S_{obs} plus the empirical residuals to get the adjusted sample

$$\theta_i^a = \hat{\mu}(S_{obs}) + \hat{\sigma}(S_{obs})\hat{\sigma}(S_i)^{-1}(\theta_i - \hat{\mu}(S_i).$$

- Extensions with localization and weighting, multivariate θ.

Further refinements - regression

Beaumont *et al.*, 2002, Blum *et al.*, 2010



Bayesian predictive model checking

- One way to perform Bayesian model checking is via a Bayesian p-value.
- Define model checking statistic as $D(y)$, p-value defined as

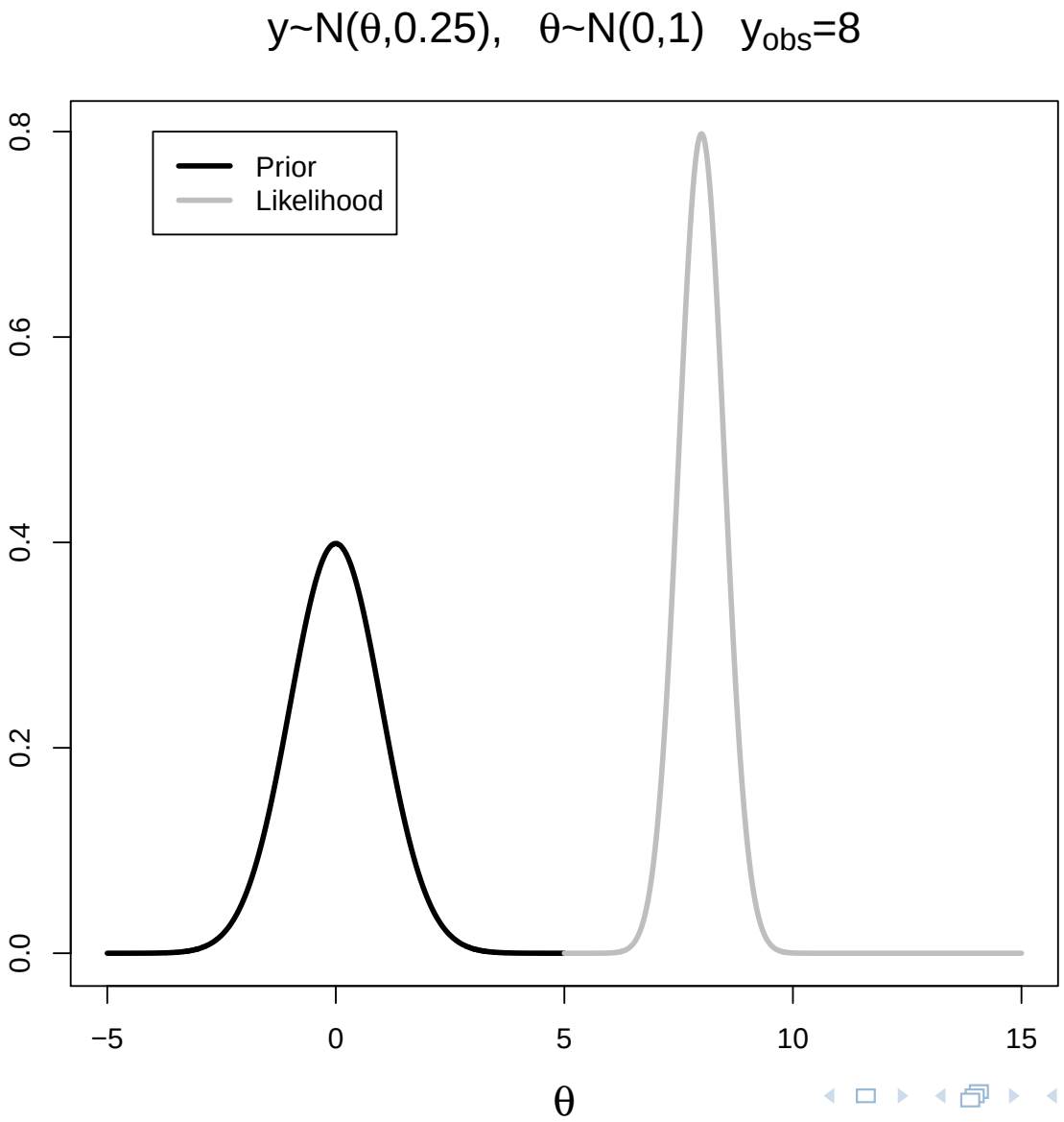
$$p = P(D(y) \geq D(y_{obs})),$$

  where $y \sim r(y)$ where $r(y)$ is some reference distribution.
- E.g. $r(y) \equiv p(y) = \int p(\theta)p(y|\theta)d\theta$, prior predictive p-value.
- Small p-value indicates a lack of model fit.

Weakly informative priors

- A value λ_0 is tentatively chosen on subjective grounds (the "base prior").
- We may want to choose a λ such that $p(\theta|\lambda)$ is weakly informative compared to $p(\theta|\lambda_0)$.
- Evans and Jang (2010) consider prior-data conflict checks as a way to formalize "weakly informative", inspired by Gelman (2006).
- Such priors may be particularly useful when there is a prior-data conflict with $p(\theta|\lambda_0)$ and a sensitivity analysis is wanted.

Prior-data conflict



Prior-data conflict checking

- Evans and Moshonov (2006) suggest prior-data conflict checks can be done via prior predictive p-values with $D(y) = 1/p(S|\lambda)$ where S is a sufficient statistic.
- In practice, S could be the MLE (asymptotically sufficient).
- Evans and Jang (2010) consider distributions of prior predictive p-values for prior-data conflict checks as a way of measuring how informative one prior is relative to another.

ABC Computational approach

Need to approximate $p(S|\lambda)$ for many different S and λ. Use ABC regression methods.

Let $(\lambda_i, \theta_i, S_i)$, $i = 1, \ldots, n$ be a sample from $p(\lambda)p(\theta|\lambda)p(S|\theta)$. Fit regression model:

$$S_i = \mu(\lambda_i) + \sigma(\lambda_i)\epsilon_i.$$

Then for any λ an approximate sample from $p(S|\lambda)$ can be obtained as

$$S_i^a(\lambda) = \hat{\mu}(\lambda) + \hat{\sigma}(\lambda)\hat{\sigma}(\lambda_i)^{-1}(S_i - \hat{\mu}(\hat{\lambda}_i)),\ i = 1, \ldots, n$$

Use kernel estimate based on $S_i(\lambda)$ to approximate $p(S|\lambda)$.

ABC Computational Approach (Cont...)

Then for a sample $S_1^0, \ldots, S_n^0$ from the prior predictive under base prior, approximate conflict p-values for $p(\theta|\lambda)$ by

$$\hat{P}(S_j^0, \lambda) = n^{-1} \sum_{i=1}^{n} I(\hat{p}(S_i^a(\lambda)|\lambda) \leq \hat{p}(S_j^0|\lambda)), \; j = 1, \ldots, n.$$

$\hat{P}(S_1^0, \lambda), \ldots, \hat{P}(S_n^0, \lambda)$ generates distribution of p-values for given λ.

Weak informativity at level γ means that the γ quantile of p-value distribution for $p(\theta|\lambda)$ is greater than the corresponding γ quantile for $p_B(\theta)$.

Logistic regression example

Racine *et al.*, 1986

- Dose response modelling (bioassay data of Racine *et al.*, 1986).
- 5 animals at each of 4 dose levels exposed to a toxin and number dead recorded.
- Doses $x_1 < x_2 < x_3 < x_4$ (the x's have been transformed to log scale and centred and scaled to have variance one).
- Response at dose i, $y_i \sim \text{Bin}(5, p_i)$, $\text{logit}(p_i) = \beta_0 + \beta_1 x_i$.
- Prior $\beta_0 \sim N(0, \sigma_0^2)$, $\beta_1 \sim N(0, \sigma_1^2)$.
- Base prior: $\sigma_0 = 10$, $\sigma_1 = 2.5$. Alternative priors: $\lambda = (\sigma_0, \sigma_1) \in [0.1, 10] \times [0.1, 20]$.

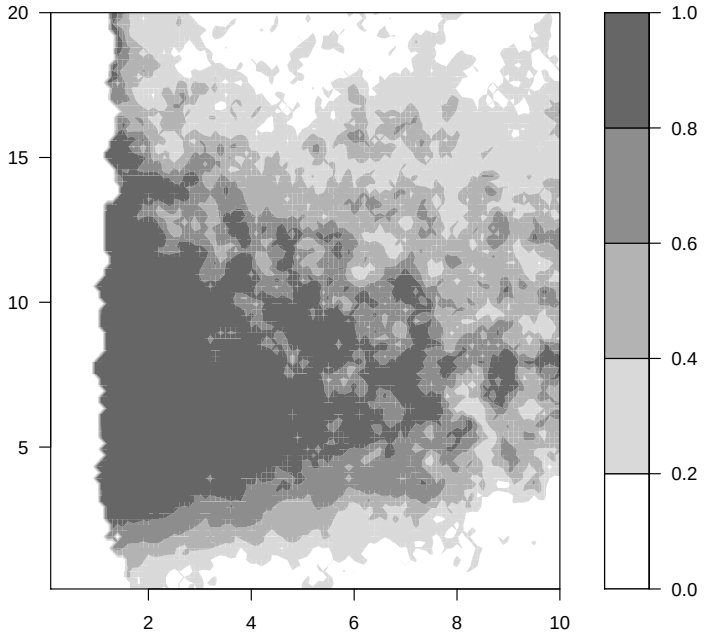
Logistic regression example

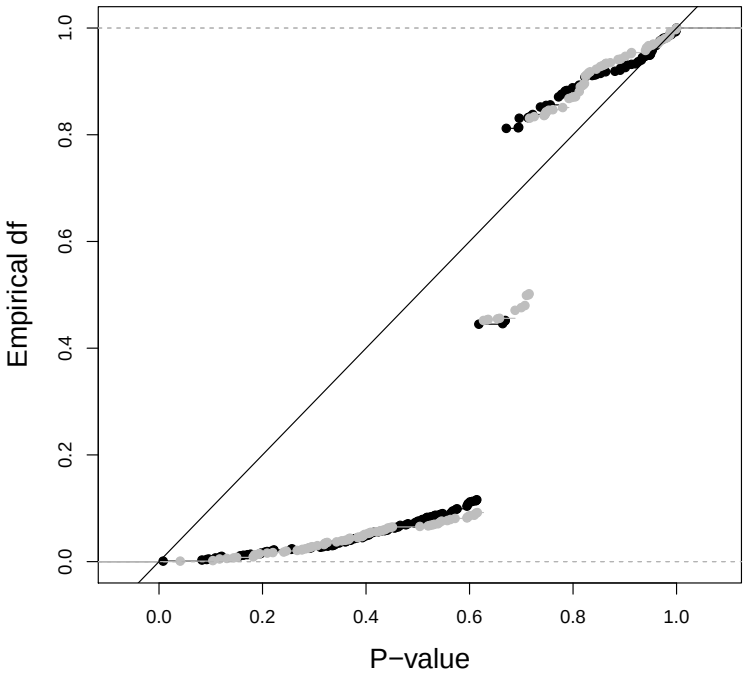
Racine *et al.*, 1986

- Use the MLE as an (asymptotically) sufficient statistic.
- Look at degree of weak informativity of the alternative prior versus the base prior at level 0.05 as λ varies.

Logistic regression example

Racine *et al.*, 1986





Can we use model checking for elicitation?

- Often the information you want to put into a prior may fall far short of determining the prior uniquely.
- Consider prior-data conflict checks for hypothetical data as providing constraints.
- For summary statistics $S^1, \ldots, S^k$, certain values $S_0^1, \ldots S_0^k$ should be considered as either "surprising" or "unsurprising" where that means failing or passing a prior-data conflict check.
- Look at how prior-data conflict p-values change with hyperparameters to inform prior choice.

Logistic regression example again

Racine *et al.*, 1986

- In this example consider Cauchy priors on β_0 and β_1 with location zero and scale λ_1 and λ_2 respectively.
- Consider using ABC to approximate how a prior-data conflict *p*-value changes as $\lambda = (\lambda_1, \lambda_2)$ changes.

Logistic regression example again

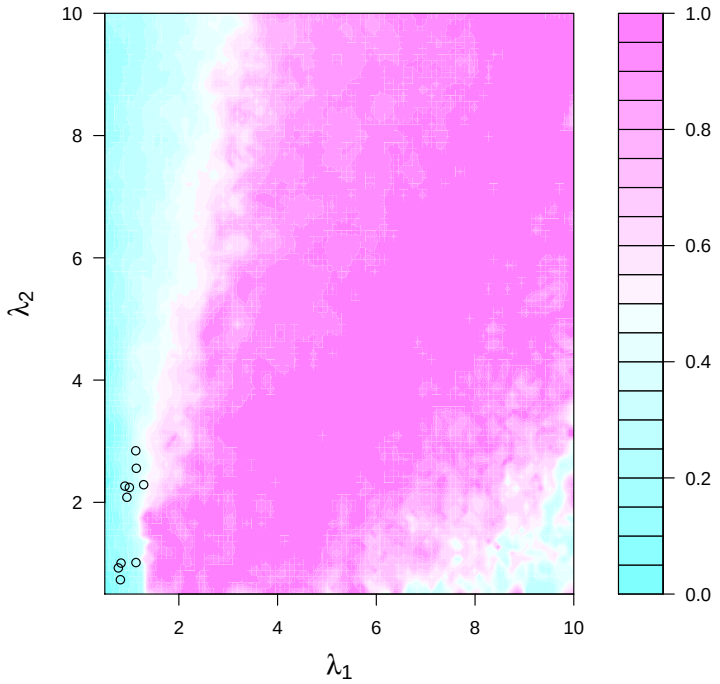
Racine *et al.*, 1986

- Conflict check: summary statistic S that is the sum of the variances of the binomial responses for the fitted model based on the MLE (or some approximation).
- Conflict p-values for hypothetical observations $S = 0.198$ and $S = 1.974$.
- We want a "reasonable" λ to:
  - fail the conflict check for $S = 0.198$ (the value of S that would result from fitted probabilities of 0.01 or 0.99 at each covariate)
  - pass the check for 1.974.
- A default prior suggested in the literature (Gelman *et al.*, 2008) has $\lambda_1 = 10$ and $\lambda_2 = 2.5$.

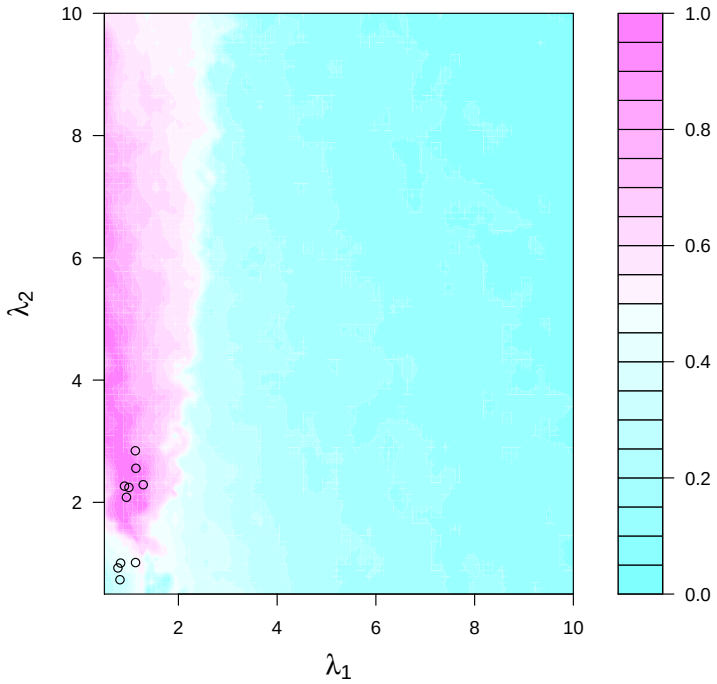
Logistic regression example again

Racine *et al.*, 1986

$S = 0.198$



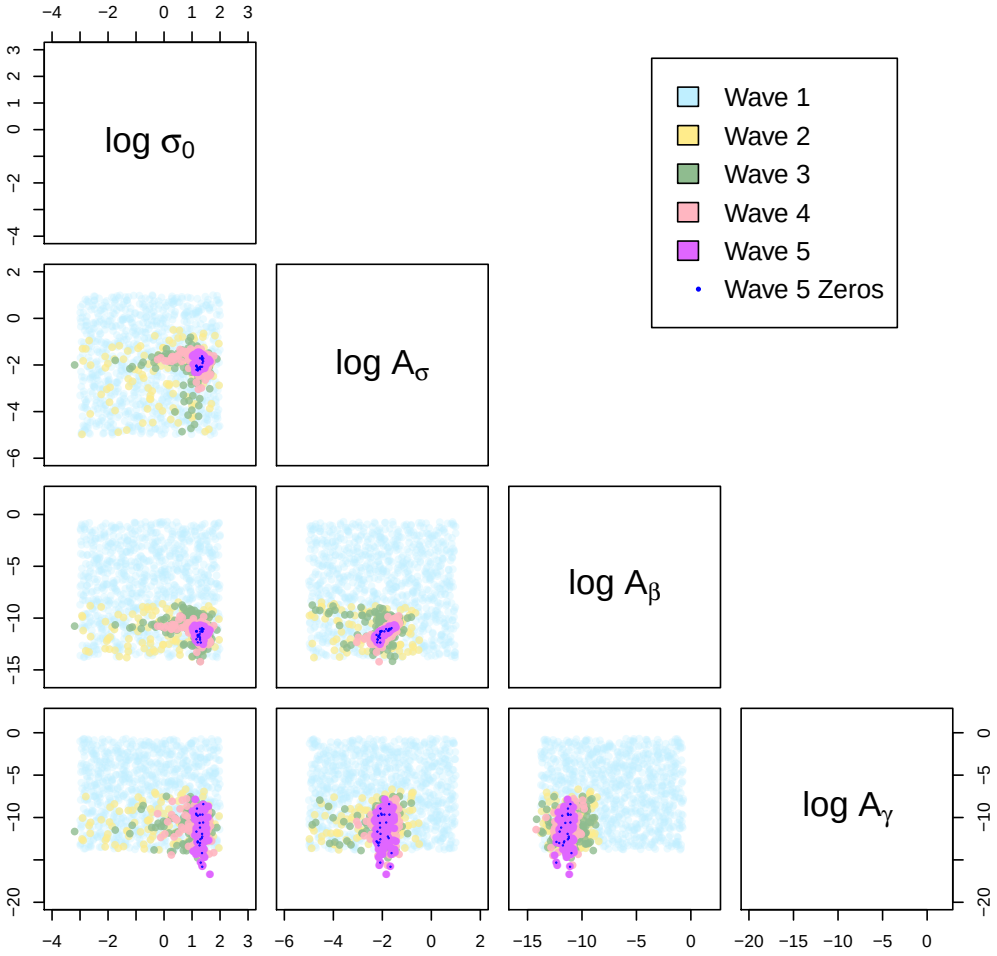
$S = 1.974$



Model checks for elicitation

- When the number of hyperparameters and model checks is large then it is not feasible to plot conflict p-values over a grid of hyperparameter values.
- In our paper we use a method called history matching to efficiently explore hyperparameter space.
- In the paper we consider an example with 4 hyperparameters.

Example: sparse signal shrinkage priors



Conclusion

- Regression ABC methods are very useful for certain repeated calculations such as those arising in
  - Prior elicitation or choice of a weakly informative prior
  - Calibration of predictive p-values under the prior
- We haven't assumed that the likelihood is intractable but the methods described might be useful in that case.

References

Nott et al (2018). Approximation of Bayesian Predictive p-Values with Regression ABC. Bayesian Analysis 13, 59-83.

Wang et al (2018). Using history matching for prior choice. To appear in Technometrics.

Beaumont et al (2002). Approximate Bayesian computation in population genetics. Genetics 162, 2025-2035.

Blum and François (2010). Non-linear regression models for approximate Bayesian computation. Statistics and Computing 20, 63-75.

Evans and Moshonov (2006). Checking for prior-data conflict. Bayesian Analysis 1, 893-914.

Evans and Jang (2011). Weak informativity and the information in one prior relative to another. Statistical Science 26, 423-439.