Parameter Estimation with Partial Information

SAND2013-4117C

H. Najm,
R. Berry, K. Chowdhary, C. Safta,
K. Sargsyan, & B. Debusschere

hnnajm@sandia.gov

Sandia National Laboratories,
Livermore, CA

Workshop on Numerical Methods for Uncertainty Quantification
Hausdorff Research Institute for Mathematics
May 13-17, 2013
Bonn, Germany

Acknowledgement

This work was supported by:

- DOE ASCR Applied Mathematics program.
- DOE Office of Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC)
- DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94-AL85000.

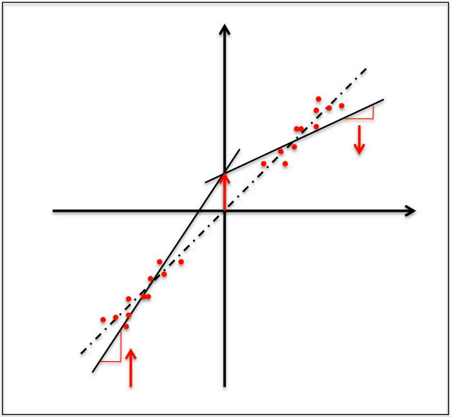
Outline

Uncertainty in Physical Model Inputs

- Probabilistic UQ requires specification of uncertain inputs
- Require joint PDF on input space
- PDF can be found given data
  - Data is rarely published
- Typically such PDFs are not available from the literature
  - What is typically available are summaries, *e.g.* nominals and bounds on data, parameters, or fitted model
  - Processed data products
- Uncertainty in computational predictions can depend strongly on detailed structure of the parametric PDF
- Need a procedure to reconstruct a PDF consistent with available information in the absence of the raw data
  - Published summaries
  - Fitting model
  - Other known experimental details

Remark – Subjective Nature of Uncertainty

- Correlation in uncertain parameters depends strongly on experimental details
- There is no intrinsic objective correlation
- Nominal values are more robust



Data range determines correlation in uncertain slope & intercept, for the same underlying truth

Bayes formula for Parameter Inference

- Data Model (fit model + noise): $y = f(\lambda) + \epsilon$
- Bayes Formula:

$$p(\lambda, y) = p(\lambda|y)p(y) = p(y|\lambda)p(\lambda)$$

$$\underbrace{p(\lambda|y)}_{\text{Posterior}} = \frac{\overbrace{p(y|\lambda)}^{\text{Likelihood}}\ \overbrace{p(\lambda)}^{\text{Prior}}}{\underbrace{p(y)}_{\text{Evidence}}}$$

- Prior: knowledge of λ prior to data
- Likelihood: forward model and measurement noise
- Posterior: combines information from prior and data
- Evidence: normalizing constant for present context

Inference with Partial/Missing Data – Scenarios

- Partial data: some of the data is missing
  - *e.g.* surveys, census, *etc*

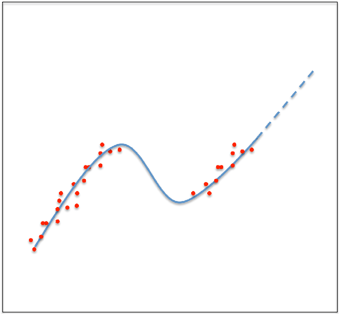
    $\Rightarrow$ Imputation methods
- All the data is missing but some information is known
  - Moments/statistics, or summaries
    - on missing data
    - on parameters fitted to missing data
    - on outputs of models fitted to missing data
  - Physical constraints
  - Details about the experiment

    $\Rightarrow$ Maximum Entropy principle

Dealing with Missing Data – Imputation

- Use information
  - from existing data
  - from prior knowledge about the model/system

  to estimate the missing data
- Interpolation is less risky than extrapolation
  - still risky
- Numerous methods for interpolation and extrapolation
- Assumptions about the structure of the model in the no-data regions



Inference with Partial Data – Imputation

Bayesian Multiple Imputation (Rubin, 1987)

- Use the posterior predictive conditioned on *observed* data to generate replicates of the missing data z_{mis}
- For each full data set, $(z_{\text{obs}}, z_{\text{mis}})$ apply Bayesian inference to get a posterior density $p(\beta | z_{\text{obs}}, z_{\text{mis}})$
- Marginalize over the missing data to get the observed data posterior $p(\beta | z_{\text{obs}})$

$$p(\beta | z_{\text{obs}}) = \int q(\beta | z_{\text{obs}}, z_{\text{mis}}) f(z_{\text{mis}} | z_{\text{obs}}) \, \mathrm{d}z_{\text{mis}}$$

No Data – Inference based on Constraints

Maximum Entropy (MaxEnt) Principle:

Maximize ignorance while satisfying given constraints $\mathcal{C}$

Given data z, use Bayes rule to get

$$q(\beta|z) = \frac{q(z|\beta)q(\beta)}{q(z)}$$

When data z is unavailable, but its distribution $w(z|\mathcal{C})$ can be estimated, then, given a prior $q(z, \beta)$, maximizing relative entropy $\mathcal{E}(p, q)$ leads to the joint MaxEnt posterior

$$p(z, \beta|\mathcal{C}) = q(\beta|z)w(z|\mathcal{C})$$

Marginalizing over z: $\quad p(\beta|\mathcal{C}) = \int q(\beta|z)w(z|\mathcal{C})\mathrm{d}z$

Data Free Inference (DFI) (Berry *et al.*, JCP 2012)

- **Intuition**: In the absence of data,
  - the structure of the fit model
  - nominals and bounds
  - physical constraints
  - other known experimental details

  implicitly inform the correlation between the parameters
- **Goal**: Make this information *explicit* in the joint PDF
- **DFI**: discover a consensus joint PDF on the parameters consistent with given information
  - Maximum entropy principle
  - Approximate Bayesian Computation (ABC) methods

DFI Algorithm in General

Given fitted model $f(x,\beta)$ and statistics from missing data z^*

$z^* \Rightarrow p(\beta|z^*) \Rightarrow S(G(p(\beta|z^*))) \Rightarrow \mathcal{S}(z^*) := \mathcal{S}^*$

- MCMC chain on data space
  - Each chain step $\Rightarrow$ data z
  - Likelihood $\pi_\epsilon(z) = K_\epsilon(|\mathcal{S}(z) - \mathcal{S}^*|)$
    - $\epsilon \to 0 \Rightarrow \{\pi_\epsilon(z) \neq 0 \Leftrightarrow \mathcal{S}(z) \to \mathcal{S}^*\}$
    - $\mathcal{S}(z)$ evaluated from the inner MCMC chain on the parameter space
- MCMC chain on parameter space
  - Estimate posterior $p(\beta|z)$
  - Evaluate statistic $\mathcal{S}(z)$
- Pooling/marginalizing consistent posteriors $p(\beta|z)$

Opinion Pooling – Linear

Consider M experts, each providing an opinion (density) $p_i(\beta)$.
Seek a consensus pooled density p_0

$$p_0 = T(p_1, \ldots, p_M)$$

Linear average pooling

$$p_0(\beta) = \int_\Delta p(\beta|z) w(z) dz$$

– Marginalization before or after pooling is equivalent

With z sampled from $w(z)$,

$$p_0(\beta|z_1, \ldots, z_M) = \frac{1}{M} \sum_{i=1}^{M} p(\beta|z_i)$$

Opinion Pooling – Logarithmic

Logarithmic average pooling

$$\ln p_0(\beta) = C + \int_\Delta \ln p(\beta|z) w(z) dz$$

- Externally Bayesian; commutes with Bayesian updating
- Does not satisfy the marginalization property

With z sampled from the density $w(z)$

$$p_0(\beta|z_1, \ldots, z_M) = \left[\prod_{i=1}^{M} p(\beta|z_i)\right]^{1/M}$$

DFI Demonstration – Plan

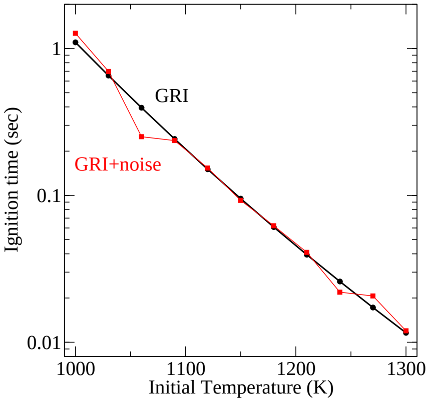
Context: Chemical ignition in a homogeneous CH_4-Air system

- Use simulated noisy data to infer a joint posterior PDF on parameters of a chemical model
- Discard data, retaining summary information
- Reconstruct the posterior using available information
- Compare the “true" and DFI posteriors

Generate ignition "data" using a detailed model+noise

- Ignition using a detailed chemical model for methane-air chemistry
- Ignition time versus Initial Temperature
- Multiplicative noise error model
- 11 data points:

$$\tau_i^d = \tau^{\text{GRI}}(T_i^o)\,(1+\sigma\epsilon_i)$$
$$\epsilon \sim N(0,1)$$



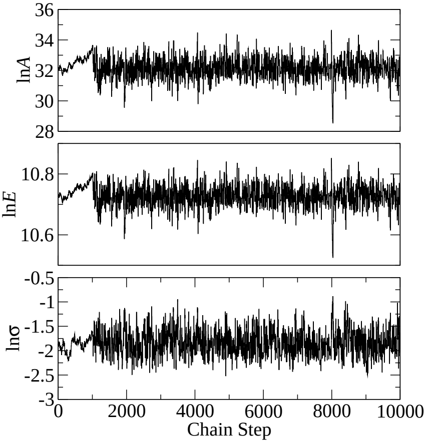
Fitting with a simple chemical model

- Fit a global single-step irreversible chemical model

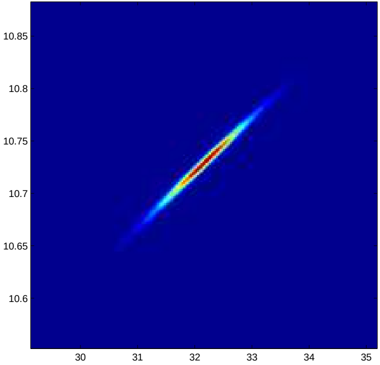
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

$$\begin{aligned} \mathfrak{R} &= [CH_4][O_2]k_f \\ k_f &= A\exp(-E/R^oT) \end{aligned}$$

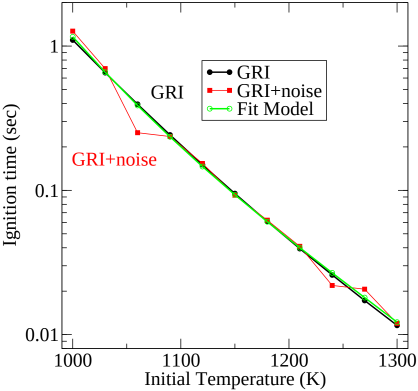
- Infer 3-D parameter vector $(\ln A, \ln E, \ln \sigma)$
- Good mixing with adaptive MCMC when start at MLE



Bayesian Inference Posterior and Nominal Prediction



Marginal joint posterior on $(\ln A, \ln E)$ exhibits strong correlation



Nominal fit model is consistent with the true model

Data Free Inference Challenge

Discarding initial data, reconstruct marginal $(\ln A, \ln E)$ posterior using the following information

- Form of fit model
- Range of initial temperature
- Nominal fit parameter values of $\ln A$ and $\ln E$
- Marginal 5% and 95% quantiles on $\ln A$ and $\ln E$

Further, for now, presume

- Multiplicative Gaussian errors
- $N = 8$ data points

DFI Algorithm Structure

- Explore the space of hypothetical data sets
  - MCMC chain on the data
  - Each state defines a data set
- For each data set:
  - MCMC chain on the parameters
  - Evaluate statistics on resulting posterior
  - Accept data set if posterior is consistent with given information
- Evaluate pooled posterior from all acceptable posteriors
  Logarithmic pooling:

$$p(\lambda|y) = \left[\prod_{i=1}^{K} p(\lambda|y_i)\right]^{1/K}$$

Computational issues

- Outer Chain:
  - N data points $\Rightarrow$ chain is $(2N+1)$–dimensional
    - Single-site update – easy to tune in hi-D
  - Markov chain on a manifold in data space
    - Ideally, use geometric MCMC methods
    - No sense of good mixing – no peak
    - Pursue convergence with # of samples
- Inner Chain:
  - Highly peaked posterior ridge
    - Adaptive MCMC for parameter estimation
    - Start at nominal model
    - Ideal proposal distrib. uses Hessian of Likelihood
- Computationally challenging:
  - Use a Polynomial Chaos surrogate for forward model
  - Run multiple outer chains in parallel, and aggregate resulting acceptable data sets

Data Chain Likelihood Function

- With β_0 being the given nominal parameter vector

$$L(z, \ln\sigma; \beta_0) = F_\delta(z)\,\frac{p(z, \ln\sigma \mid \beta_0)}{\max_{z,\ln\sigma}[p(z, \ln\sigma \mid \beta_0)]}$$

- Consistency check: Nominals

$$p(z, \ln\sigma \mid \beta_0) \propto p(z \mid \beta_0, \ln\sigma)\pi(\beta_0, \ln\sigma)$$

– parameter likelihood $\times$ prior; inner chain

- Consistency check: Bounds
  - Evaluate the marginal parameter posterior $p(\beta|z)$ quantiles
  - $F_\delta(z)$ is a Kullback-Leibler density that is maximal when the posterior quantiles are consistent with the given 5%, 95% quantiles on β

Bounds Check employing the Kullback-Leibler Density

- $[p,r,q]$: trinomial density with probabilities $p+r+q=1$
- KL density measuring mismatch between two trinomials:

$$f_\delta([p,r,q]\,|\,[0.05,0.90,0.05]) = \exp\left\{-\delta\left(p\ln\frac{p}{0.05} + r\ln\frac{r}{0.90} + q\ln\frac{q}{0.05}\right)\right\}$$

Then, with the measured $p(\beta|z)$ quantiles

$$p_1(z) = P[\,\ln A < (\ln A)_{0.05}\,|\,z\,] \qquad q_1(z) = P[\,\ln A > (\ln A)_{0.95}\,|\,z\,]$$
$$p_2(z) = P[\,\ln E < (\ln E)_{0.05}\,|\,z\,] \qquad q_2(z) = P[\,\ln E > (\ln E)_{0.95}\,|\,z\,]$$

we define

$$F_\delta(z) \propto \prod_{i=1}^{2} f_\delta\left([p_i(z), 1-p_i(z)-q_i(z), q_i(z)]\,\middle|\,[0.05,0.90,0.05]\right)$$

Inner/Parameter Chain Likelihood

Likelihood function employs, by construction, an *i.i.d.* multiplicative gaussian noise model for the dependent data τ

$$\frac{\tau}{\tau^m(T^o,\beta)} - 1 \sim N(0,\sigma^2)$$

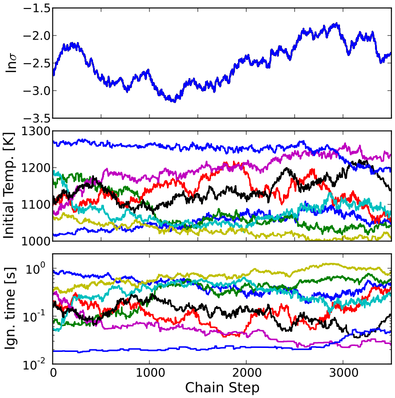
and an *i.i.d.* uniform model for the independent data T^o.

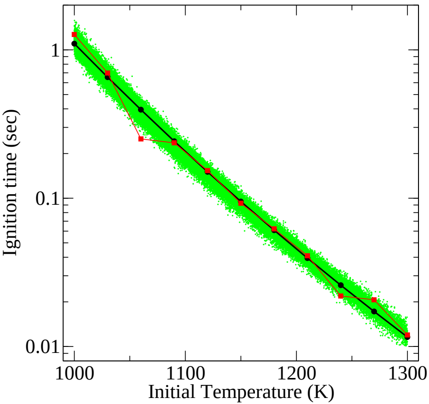
$$T^o \sim U(T^o_{\min}, T^o_{\max})$$

Giving the likelihood for $z = (\tau^d_1, \ldots, \tau^d_N, T^o_1, \ldots, T^o_N)$:

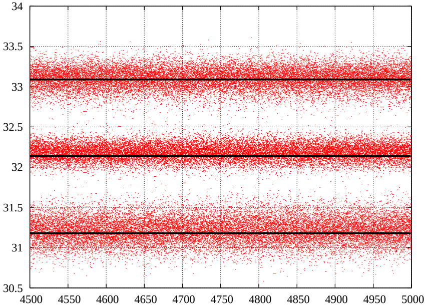
$$p(z \,|\, \beta, \ln\sigma) \propto \frac{1}{(2\pi\sigma^2)^{N/2}} \frac{1}{\prod_{i=1}^N \tau^m(T^o_i,\beta)} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N \left[\frac{\tau^d_i}{\tau^m(T^o_i,\beta)} - 1 \right]^2 \right\}$$

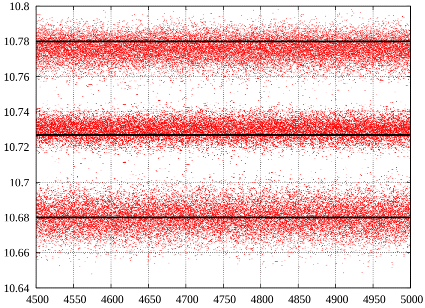
Short sample from outer/data chain





Scatter of Nominals and 0.05/0.95 Quantiles



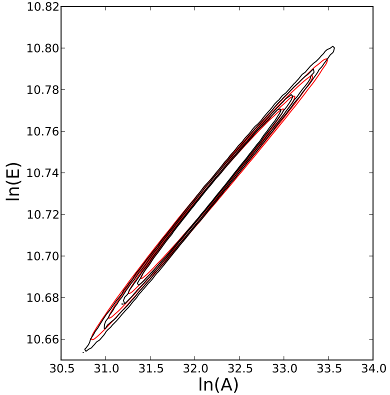


- Parameter posterior statistics are in general agreement with the desired nominal/quantile summaries
- Amplitude of scatter can be controlled by choice of weights in the likelihood function of the data chain

Pooled DFI Posterior vs. Reference Posterior

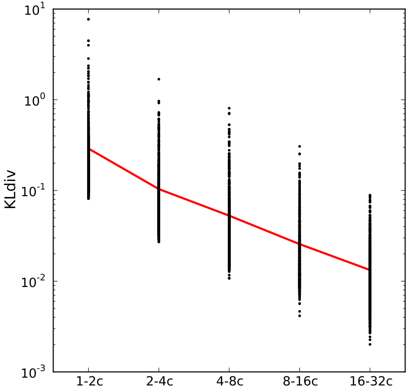
- 50 data chains run in parallel
  $\Rightarrow$ Pooled DFI posterior
- Structure of 2D pooled posterior is robust
  - Irrespective of the number of chains
- Very close similarity to the reference posterior
- Available information is essentially a sufficient statistic

175,000



Empirical Convergence of 3D Pooled Posteriors

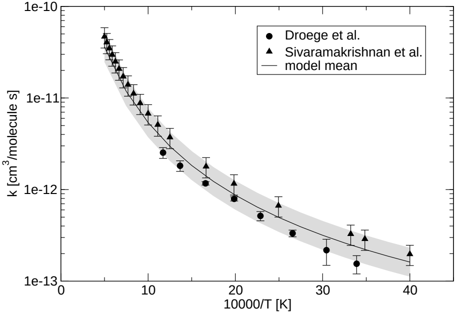
- Kullback-Leibler divergence between posteriors of successively increased data volumes
- Combinatorial choices of chains pooled at each stage
  - statistical scatter of KLdiv
- Overall convergence evident $\propto 1/N$



Processed Data Products

Frequently, we are given published “data" in the form of nominal values and error bars

- Fitting of $y = f(x; \beta)$
- Data:
  $(x_\ell, y^o_\ell \pm \alpha_\ell), \quad \ell = 1, \cdots, L$
  $\pm\alpha_\ell$ are *e.g.* $\pm 2\sigma$ error bars



Two interpretations of the (y^o_ℓ, α_ℓ) as nominals and bounds

- on the **data**: N observed random y_ℓ at each x_ℓ
- on the **fitted model**: uncertain $y_\ell = f(x_\ell, \beta)$

Processed Data Products – Demo

Demonstrate DFI on a simple model problem:

- Generate synthetic noisy data from a double exponential decay model

$$\begin{aligned} s &= 2e^{-2t} + .5e^{-.5t} \\ y &= s + (\sigma_1 s + \sigma_2)\eta \end{aligned}$$

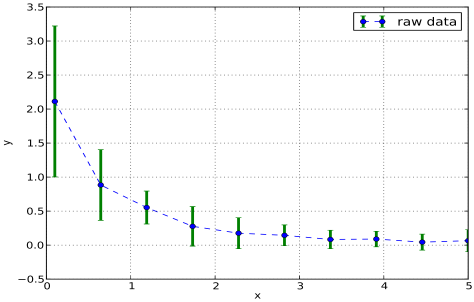
  Retain only error bars at different times

- Fit using DFI with a single exponential decay model

$$\begin{aligned} f &= a \exp(-kt) \\ y &= f + (\sigma_1 f + \sigma_2)\eta \end{aligned}$$

DFI with Processed Data Products, Case I

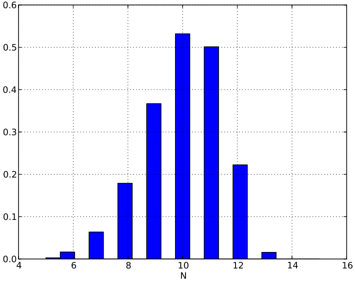
- Given means and variances on the raw data.



- Using MaxEnt: $y_\ell \sim N(y_\ell^o, \sigma_\ell^2), \quad \ell = 1, \ldots, L$
- Generate consistent data samples

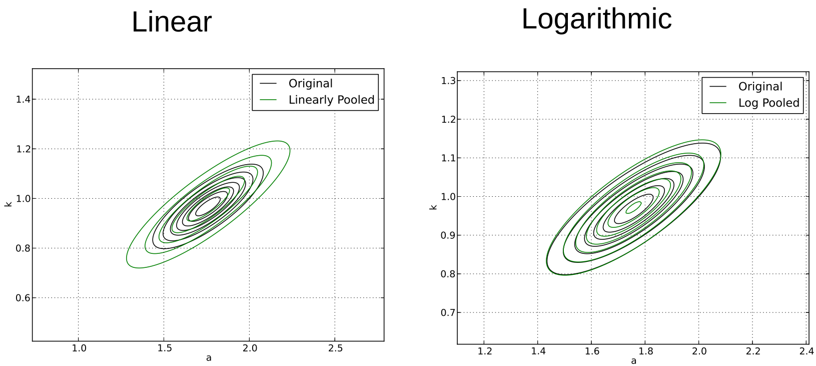
Unknown Dimensionality: Marginalizing over N

- Generate consistent data sets using MaxEnt
- The number of samples N is unknown
- Place a density on the number of samples N, and sample it along with data z



Presumed Beta density on N

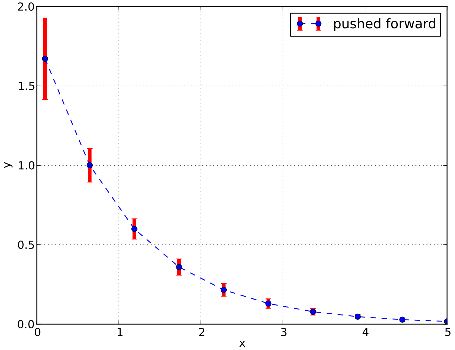
Linear and Logarithmic Pooling give Similar Results



- Pooled posteriors on (a, k), marginalized over error model hyperparameters σ_1, σ_2 and N
- Posteriors from individual data sets are largely similar to the reference posterior, hence the similar pooling results

DFI with Processed Data Products, Case II

- Given means & variances on the uncertain fitted model output



- Consider these as statistics on the pushed forward posterior
- Use these statistics to constrain proposed data sets

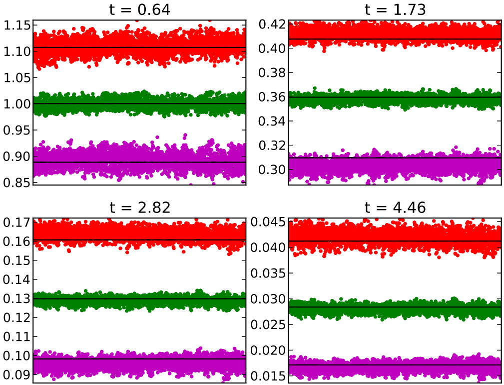
DFI Implementation

Given required statistics $\mathcal{S}^*$
Missing data z^*

$$z^* \Rightarrow p(\beta|z^*) \Rightarrow p(f(x,\beta)|z^*) \Rightarrow \mathcal{S}(z^*) := \mathcal{S}^*$$

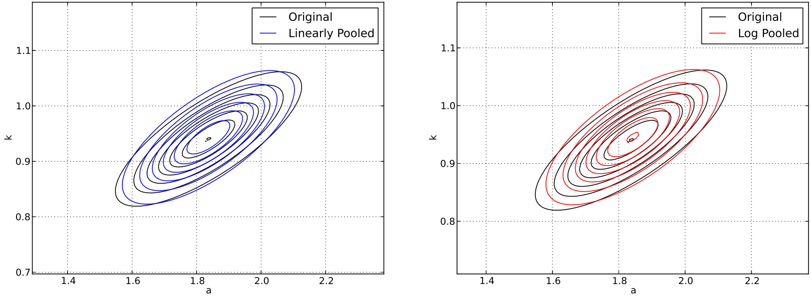
- MCMC chain on data space
  - Each chain step $\Rightarrow$ data z
  - Likelihood $\pi_\delta(z) = f_\delta(\mathcal{S}(z) - \mathcal{S}^*)$
    - $\mathcal{S}(z)$ evaluated from the inner MCMC chain on the parameter space
- MCMC chain on parameter space
  - Estimate posterior $p(\beta|z)$
  - Estimate pushed forward posterior $p(f(x,\beta)|z)$
  - Evaluate statistic $\mathcal{S}(z)$
- Pooling

PFP Quantiles



- Consistent pushed forward posterior quantiles clustered in the vicinity of reference values

Linear and Logarithmic Pooling give Similar Results



- Posteriors from individual data sets are largely similar to the reference posterior, hence the similar pooling results

Closure

Outlined several missing data contexts

- Partial data – Imputation
- No data
  - Maximum Entropy principle
  - Approximate Bayesian Computation
  - DFI
- Demonstrations
  - Constraints on fitted parameters of interest
  - Processed data products
    - Constraints on raw data
    - Constraints on fitted model uncertain predictions