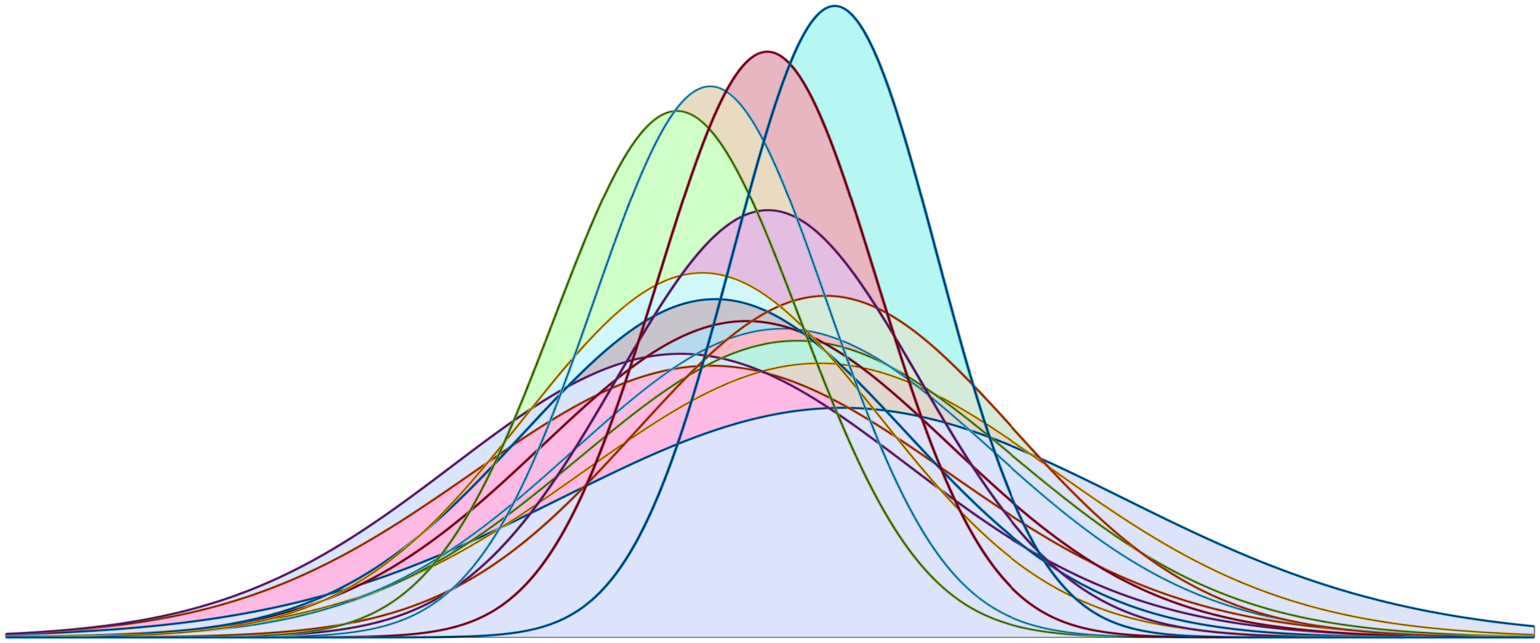
Epistemic uncertainty and Bayesian updating: Calibration



Ander Gray
Institute for Risk and Uncertainty
University of Liverpool

Contents

1. Types of uncertainty
2. Representing uncertainty
3. Reducing uncertainty:
    - Bayes’ Law and distribution updating
    - Prior Conjugation
    - Markov chain Monte-Carlo
4. Model Calibration
5. Advantages and critiques

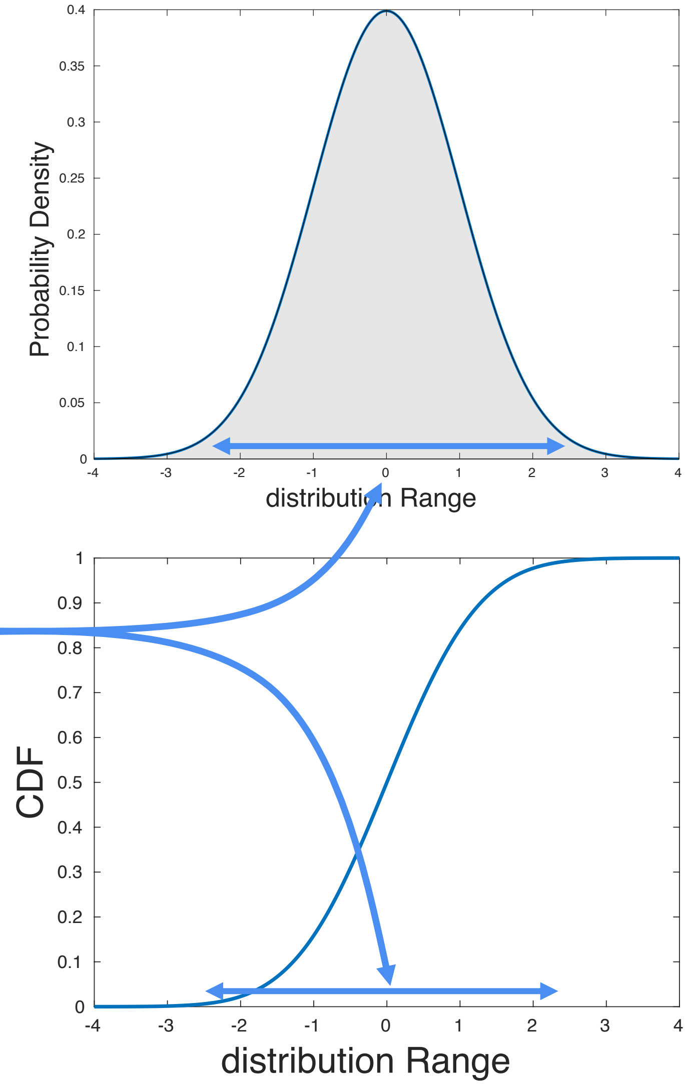
Uncertainty Characterization

1. <u>Aleatoric Uncertainty</u>:
   Randomness

- The entity is variable by it's nature
- Irreducible uncertainty
- ”Predictably unpredictable”

Examples:
- Nuclear decay
- The rolling of a dice
- Quantum mechanics

Variable has multiple true values



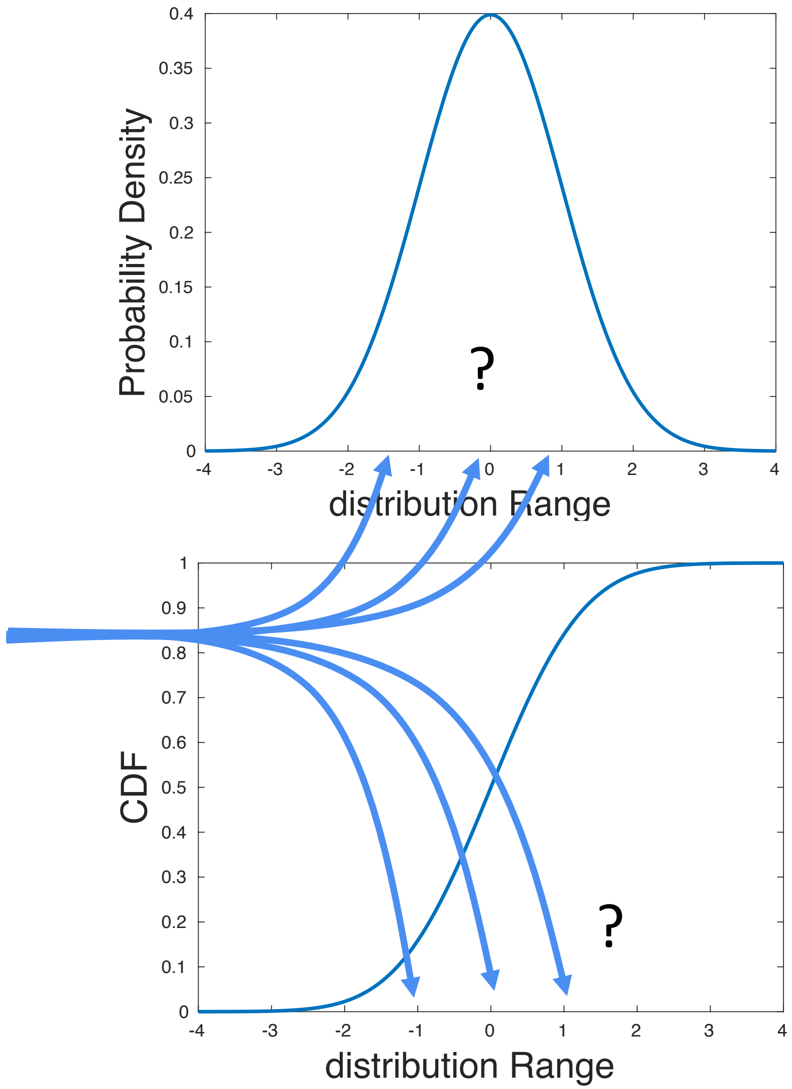
Uncertainty Characterization

2. <u>Epistemic Uncertainty</u>:
unknown

- The entity is fixed, but unknown to us
- Reducible uncertainty
- Differs between individuals: experience based

Examples:
- Expert opinion
- Ambiguity: linguistic statements
- Conflicting information
- Measurement with limited precision

Variable has fixed value which is unknown to us



Uncertainty Characterization

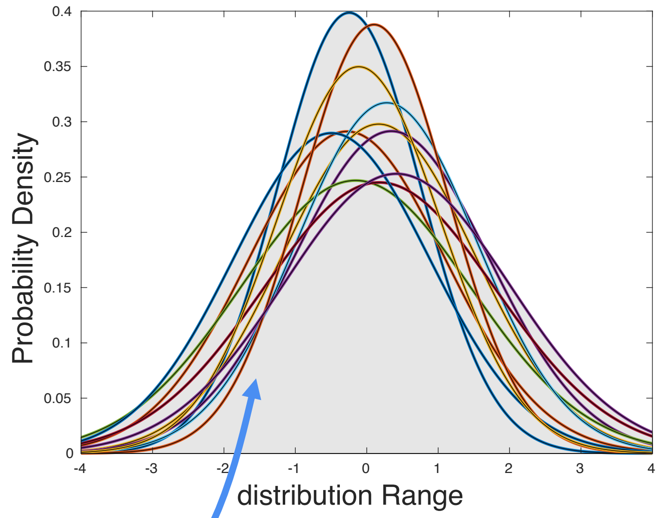
3. <u>A mixed state</u>:
uncertainty about uncertainty

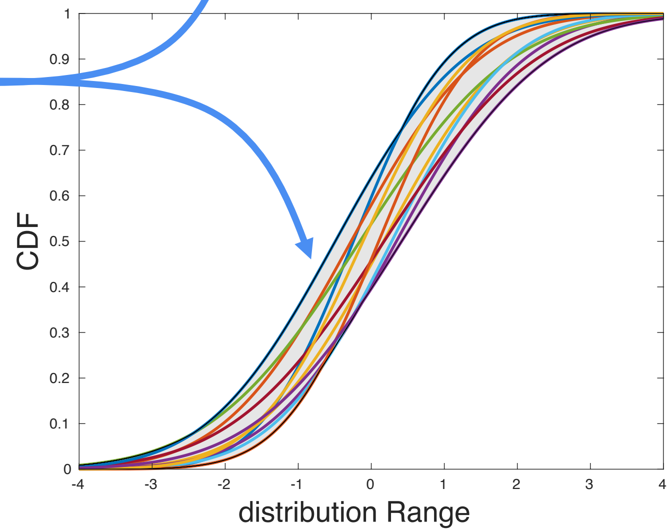
- When both aleatoric and epistemic uncertainty are present
- Epistemic uncertainty can still be reduced. Solution is exact distribution

Examples:

- Distribution parameters unknown
- Distribution shapes unknown
- Interval bounded data

Variable is aleatoric, but true distribution is unknown





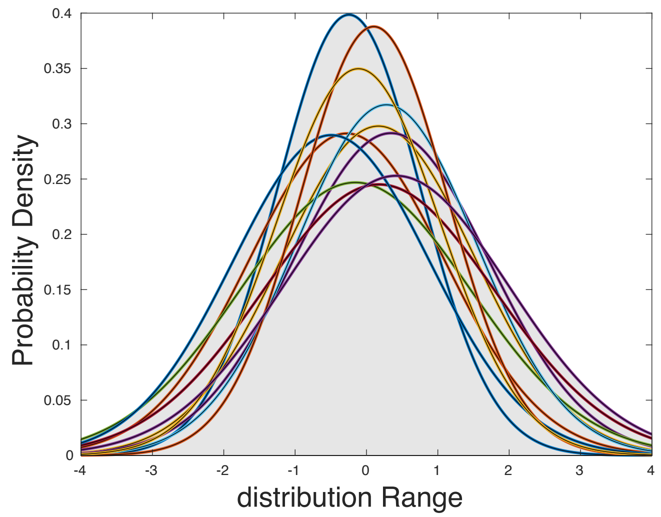
Uncertainty Characterization

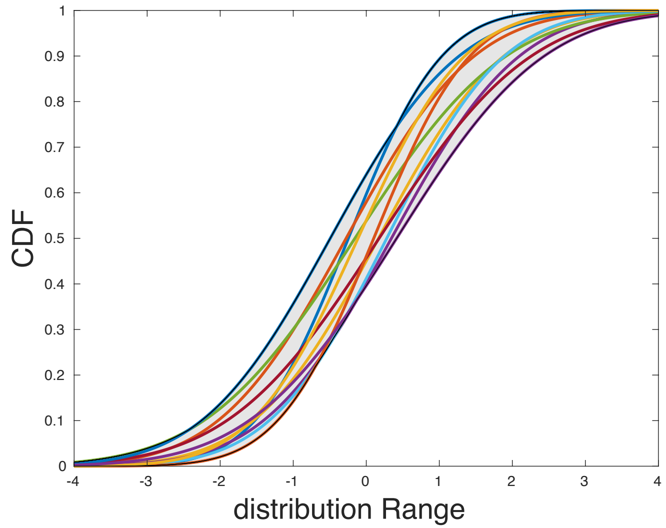
<u>Caution!</u>

- Both Epistemic and Aleatoric uncertainty should not be represented in a singular distribution

- These are two fundamentally different things, which behave differently

- Sometimes probability theory isn't enough





Uncertainty Representation

Objective probability

1. Classical probability or chance

- All events have equal probability

$$P(A) = \frac{N_A}{N}$$

2. Frequentism

- Probability is relative frequency of the event

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

Subjective probability

Bayesian

- Probability is a *degree of belief*
- Malleable using data
- You must always provide your current state of knowledge as a distribution

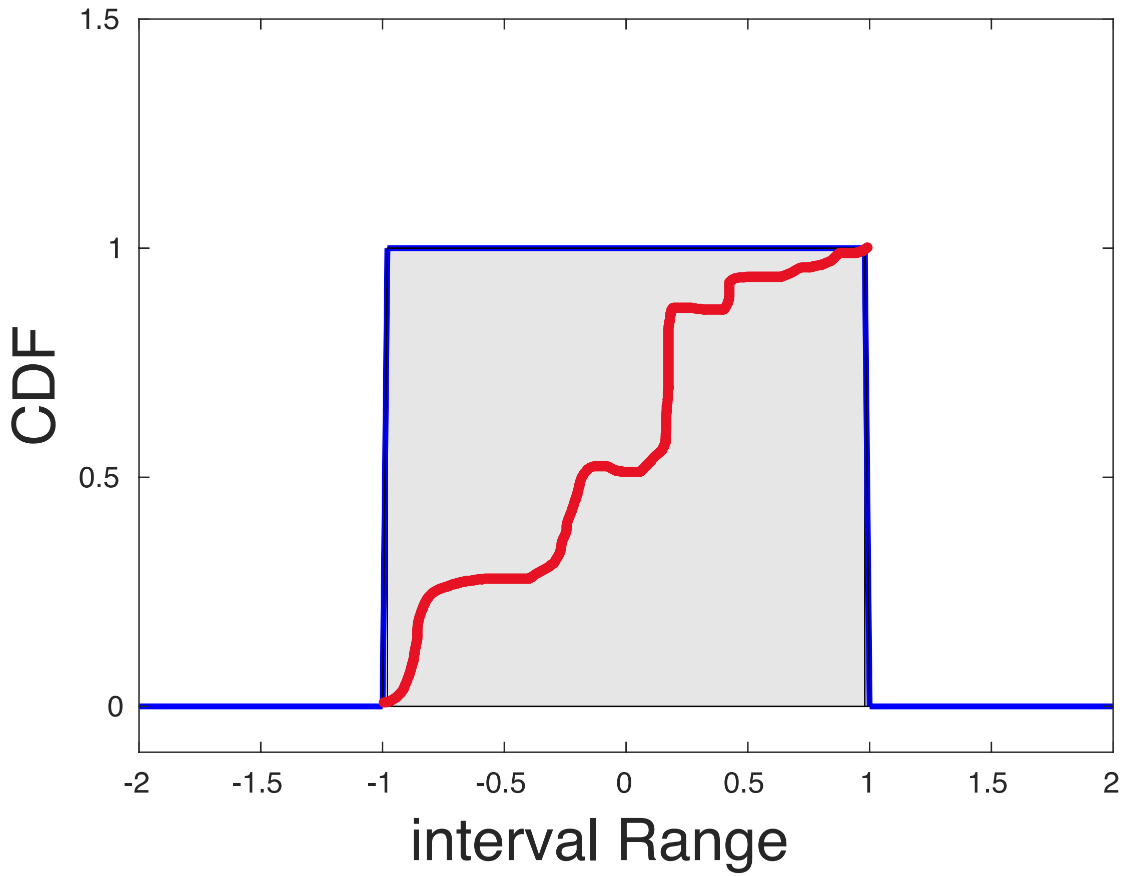
Bayesian *extremism:*

Any state of knowledge can be represented by a probability distribution

Uncertainty Representation

Non-probabilititistic methods

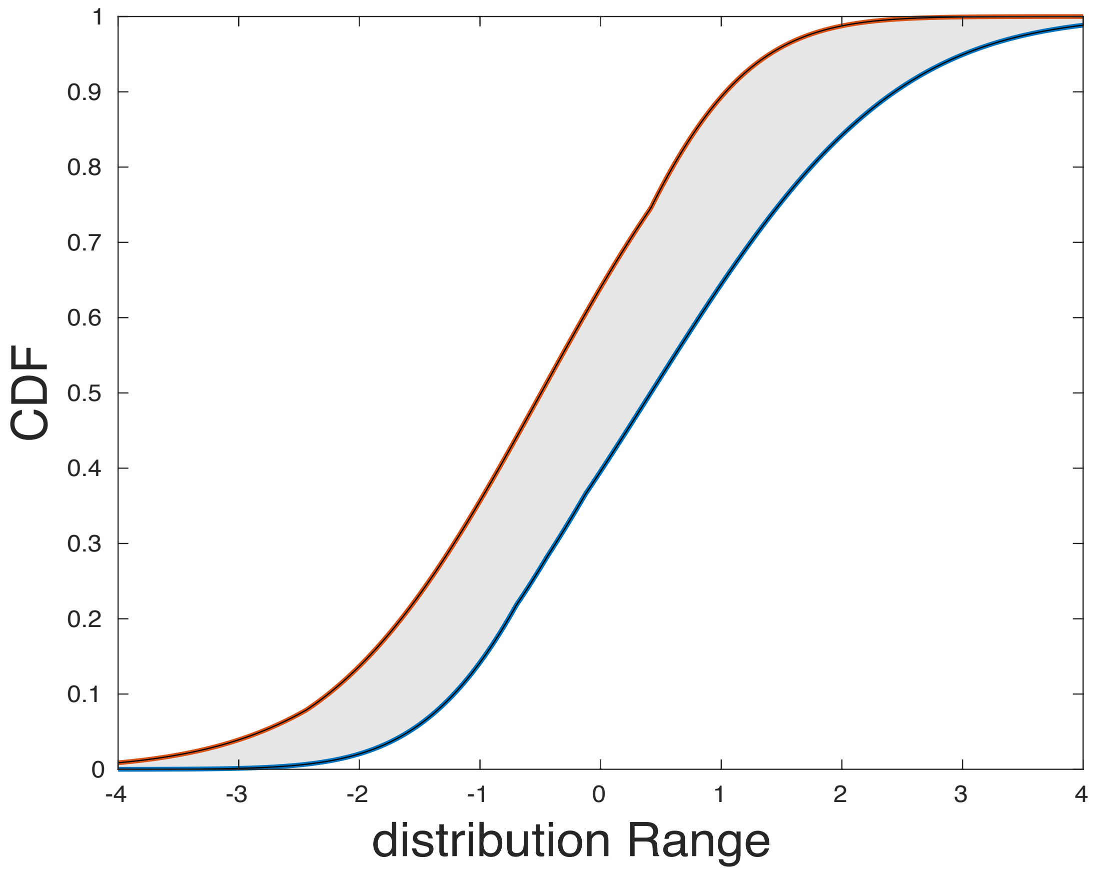
- Interval analysis



Uncertainty Representation

Non-probabilitistic methods

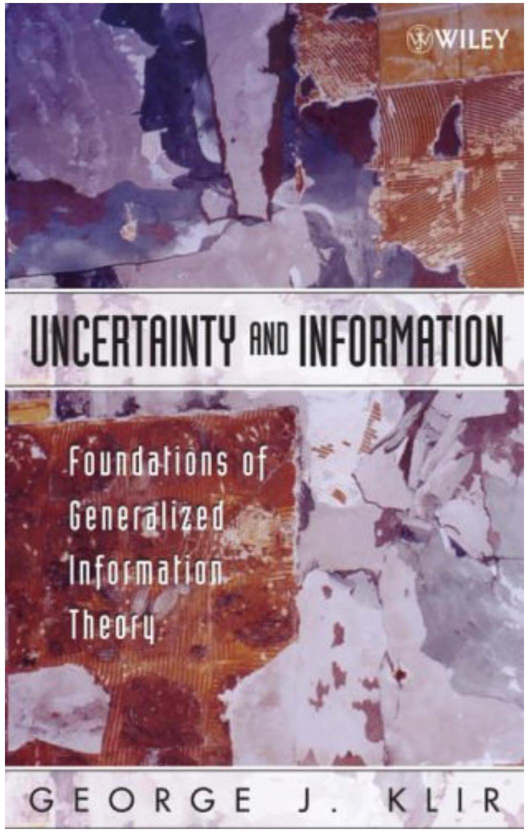
- Interval analysis
- Imprecise probability theory



Uncertainty Representation

Non-probabilitistic methods

- Interval analysis
- Imprecise probability theory
- Fuzzy sets/ Fuzzy probability theory
- Possibility theory
- Dempster-Shafer / Evidence theory

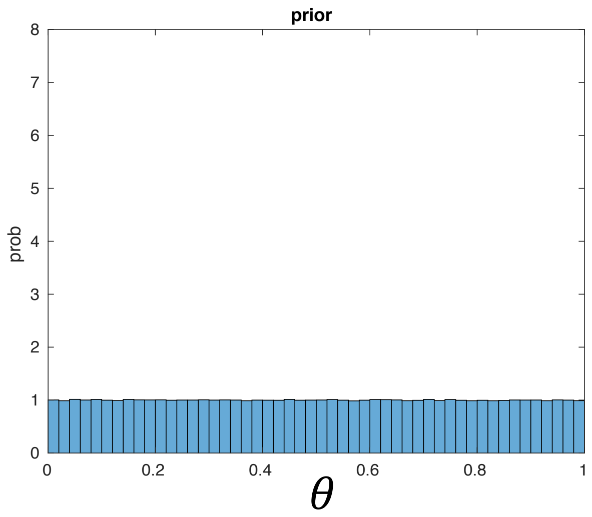


Uncertainty and Information: foundation of generalized information theory

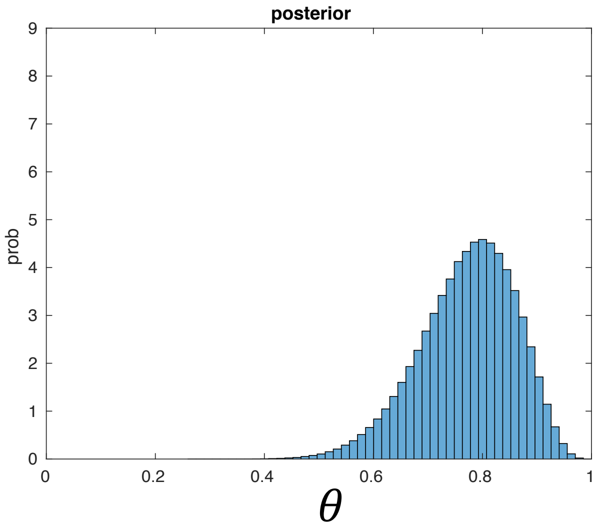
By: George J. Klir

Reducing uncertainty: Bayesian updating

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$



Data/ information about θ



Prior: $P(\theta)$

Posterior: $P(\theta|D)$

Reducing uncertainty: Bayesian updating

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Likelihood: $P(D|\theta)$

- Sometimes called the "Probabilistic model"
- Gives relative probability of every point in your sample space θ
- θ is the variable and D is fixed

Evidence: $P(D)$

$$\int_{\theta} P(D|\theta)P(\theta)d\theta$$

- The main bottleneck
- θ is integrated out, so a constant
- Numerical integration becomes too difficult at high dimensions

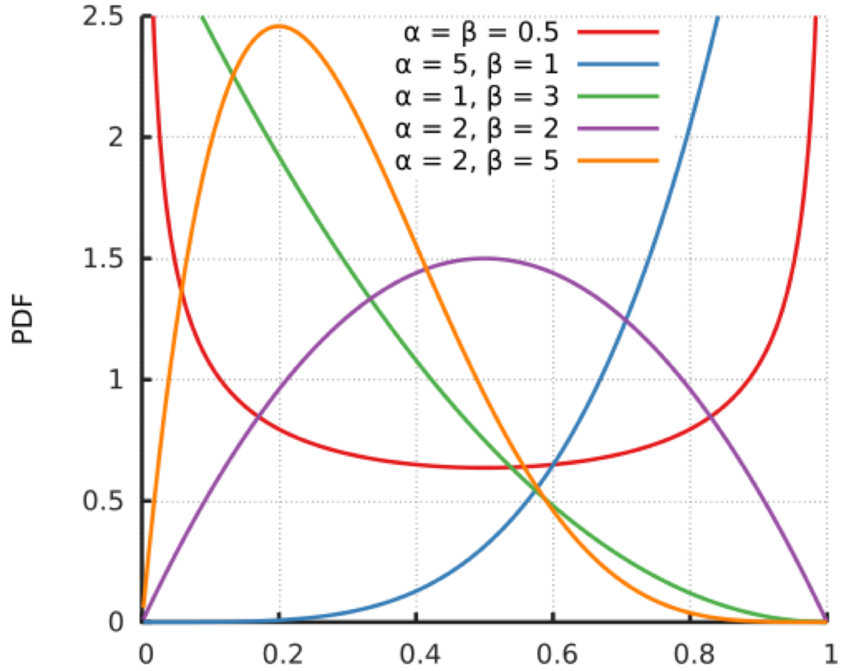
Prior conjugates

Prior and likelihood same family:

posterior	*likelihood*	*prior*
Beta	Bernoulli	Beta
Beta	Binomial	Beta
Beta	Negative Binomial	Beta
Normal	Normal	Normal
Gamma	Poisson	Gamma
Gamma	Normal	Gamma
Inverse-gamma	Exponential	Inverse-gamma

Prior conjugates

Beta prior and Binomial data:



Distribution for updating

Prior parameters: α, β

Posterior parameters

$$\alpha_{new} = \alpha + \sum D_i$$

$$\beta_{new} = \beta + N_{tot} - \sum D_i$$

D = 1,0,0,0,1,0,1,1

Binomial Data

Prior conjugates

Table of conjugate distributions [edit]

Let n denote the number of observations. In all cases below, the data is assumed to consist of n points $x_1, \ldots, x_n$ (which will be random vectors in the multivariate cases).

If the likelihood function belongs to the exponential family, then a conjugate prior exists, often also in the exponential family; see Exponential family: Conjugate distributions.

Discrete distributions [edit]

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters	Interpretation of hyperparameters[note 1]	Posterior predictive[note 2]
Bernoulli	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + n - \sum_{i=1}^{n} x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]	$p(\tilde{x} = 1) = \frac{\alpha'}{\alpha' + \beta'}$
Binomial	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]	$\operatorname{BetaBin}(\tilde{x} \vert \alpha', \beta')$ (beta-binomial)
Negative binomial with known failure number, r	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + rn$	$\alpha - 1$ total successes, $\beta - 1$ failures[note 1] (i.e., $\frac{\beta - 1}{r}$ experiments, assuming r stays fixed)	
Poisson	λ (rate)	Gamma	$k,\ \theta$	$k + \sum_{i=1}^{n} x_i,\ \frac{\theta}{n\theta + 1}$	k total occurrences in $\frac{1}{\theta}$ intervals	$\operatorname{NB}(\tilde{x} \mid k', \theta')$ (negative binomial)
			$\alpha,\ \beta$[note 3]	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + n$	α total occurrences in β intervals	$\operatorname{NB}\left(\tilde{x} \mid \alpha', \frac{1}{1 + \beta'}\right)$ (negative binomial)
Categorical	$\boldsymbol{p}$ (probability vector), k (number of categories; i.e., size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + (c_1, \ldots, c_k)$, where c_i is the number of observations in category i	$\alpha_i - 1$ occurrences of category i[note 1]	$p(\tilde{x} = i) = \frac{\alpha_i{}'}{\sum_i \alpha_i{}'} = \frac{\alpha_i + c_i}{\sum_i \alpha_i + n}$
Multinomial	$\boldsymbol{p}$ (probability vector), k (number of categories; i.e., size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + \sum_{i=1}^{n} \mathbf{x}_i$	$\alpha_i - 1$ occurrences of category i[note 1]	$\operatorname{DirMult}(\tilde{\mathbf{x}} \mid \boldsymbol{\alpha}')$ (Dirichlet-multinomial)
Hypergeometric with known total population size, N	M (number of target members)	Beta-binomial[4]	$n = N, \alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]	
Geometric	p_0 (probability)	Beta	$\alpha,\ \beta$	$\alpha + n,\ \beta + \sum_{i=1}^{n} x_i - n$	$\alpha - 1$ experiments, $\beta - 1$ total failures[note 1]	

Continuous distributions [edit]

Example

Estimate the weight of a biased coin:

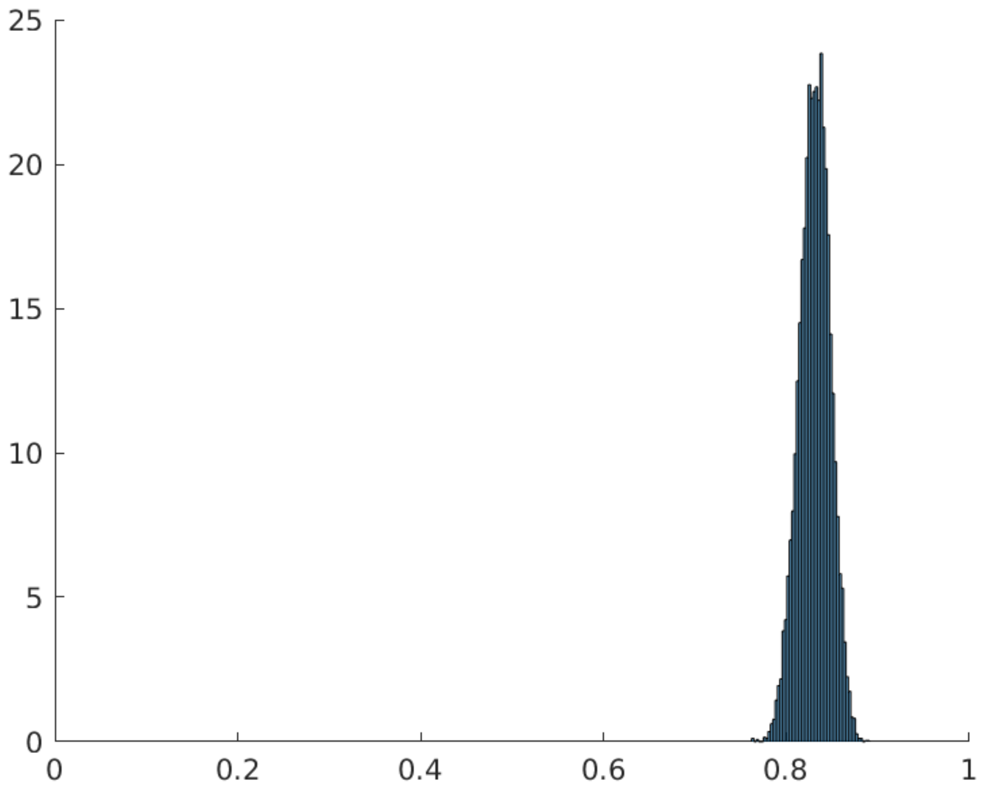
$D \sim Bernoulli(\theta)$ θ= ?

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

$N_{data} = 100$

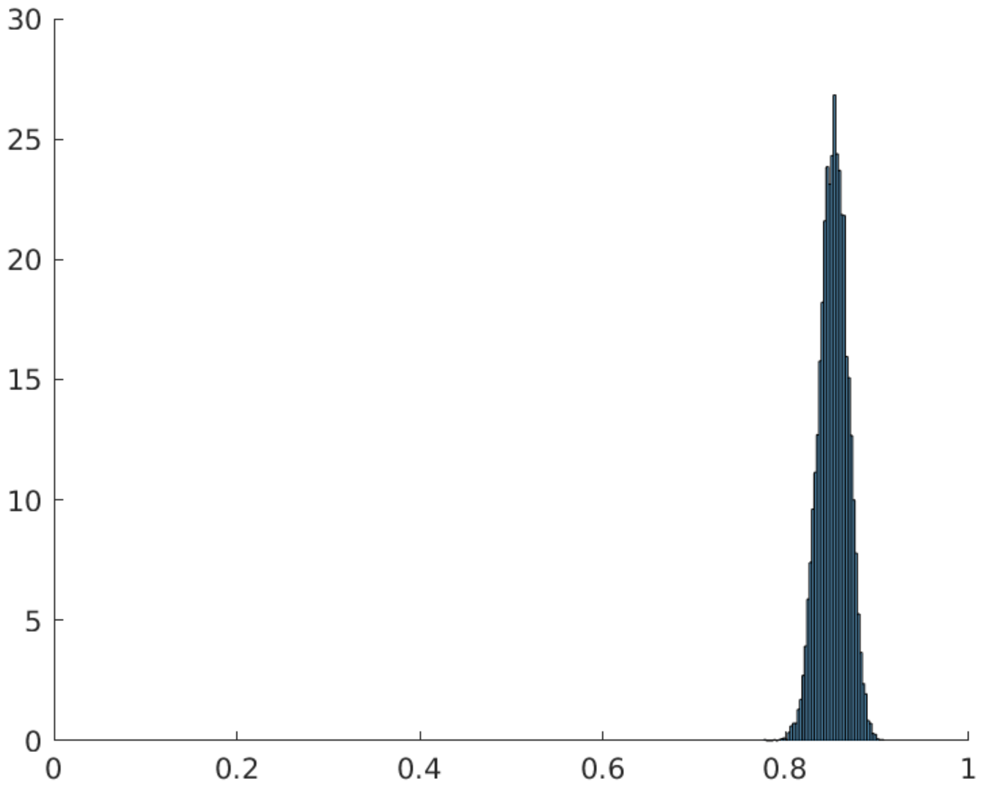
True value:

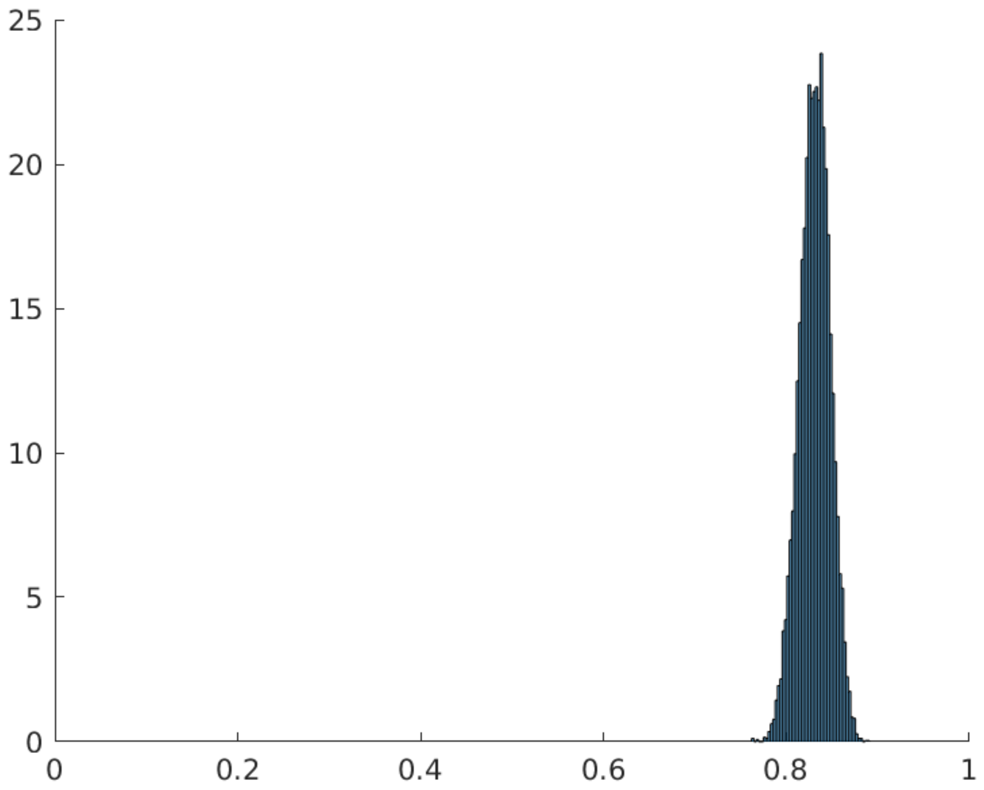
θ= 0.86



P(θ |D)

Example





$N_{data} = 100$

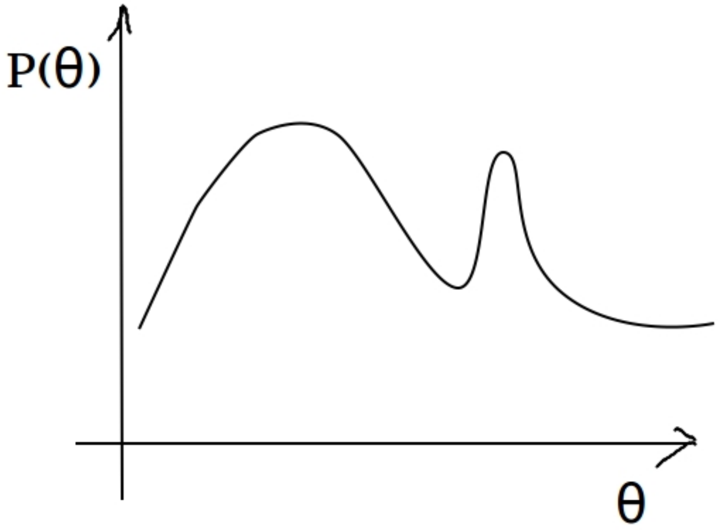
Markov-Chain Monte Carlo

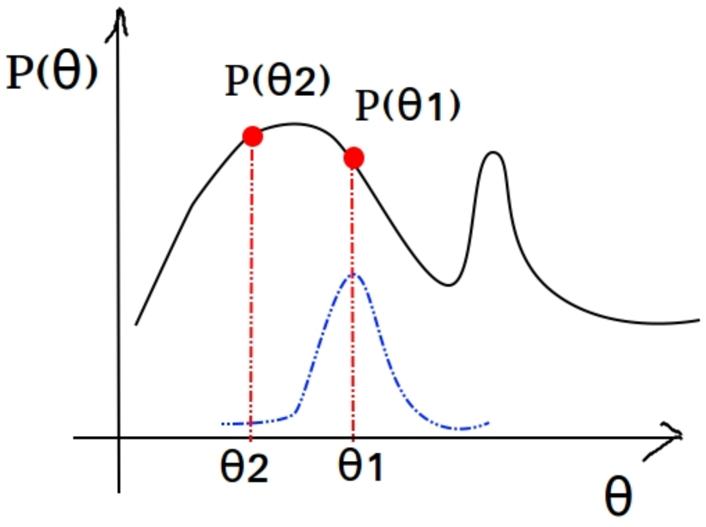
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

$$P(D) = \int_\theta P(D|\theta)P(\theta)d\theta \ = const$$

Problem boils down to sampling an unnormalized arbitrary distribution





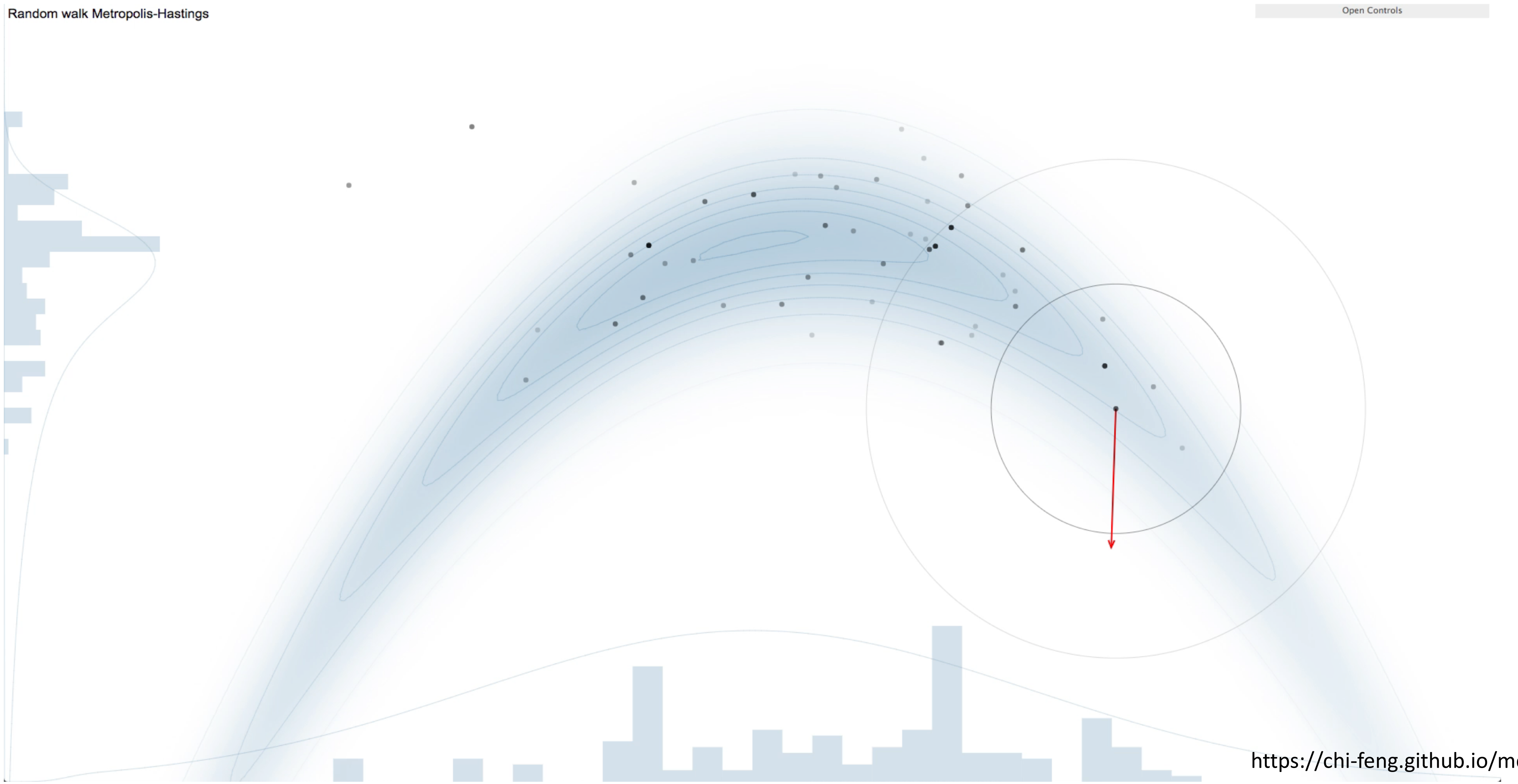
$$\alpha = \frac{P(\theta_2|D)}{P(\theta_1|D)}$$

$$x \sim U(0,1)$$

$$\alpha > \mathrm{x} \rightarrow accept$$

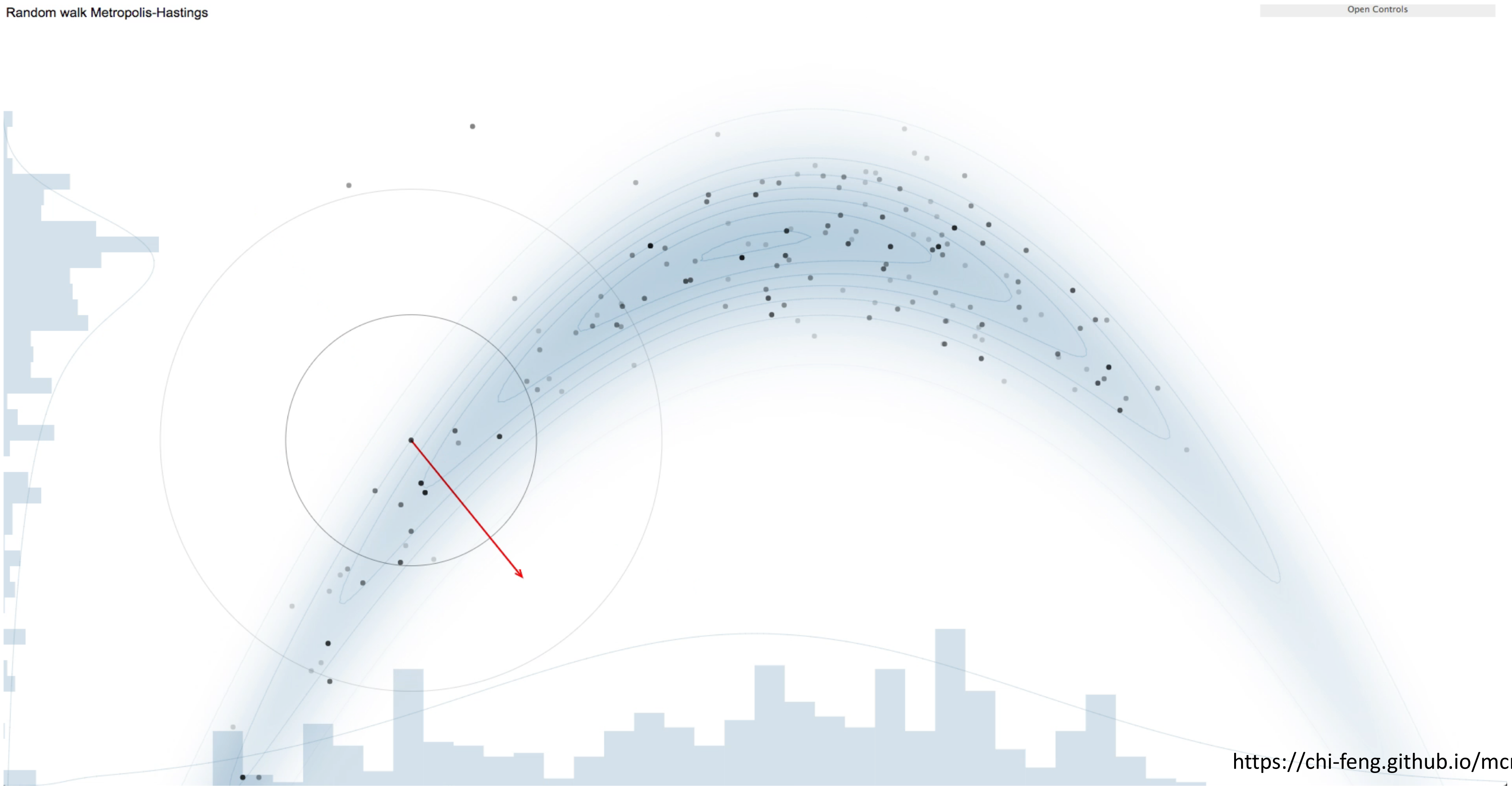
$$\alpha < \mathrm{x} \rightarrow reject$$

Markov-Chain Monte Carlo



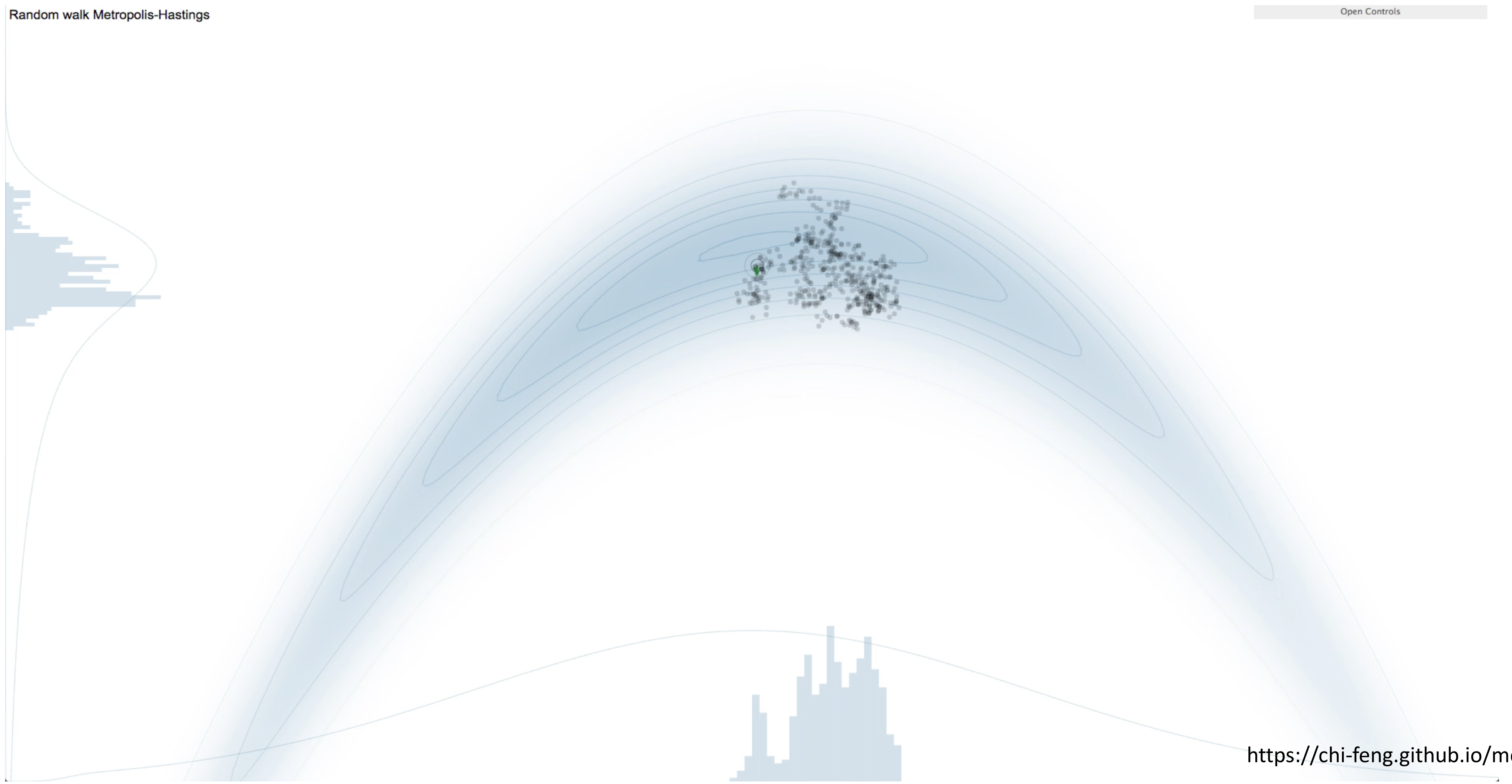
https://chi-feng.github.io/mcmc-demo/

Markov-Chain Monte Carlo



https://chi-feng.github.io/mcmc-demo/

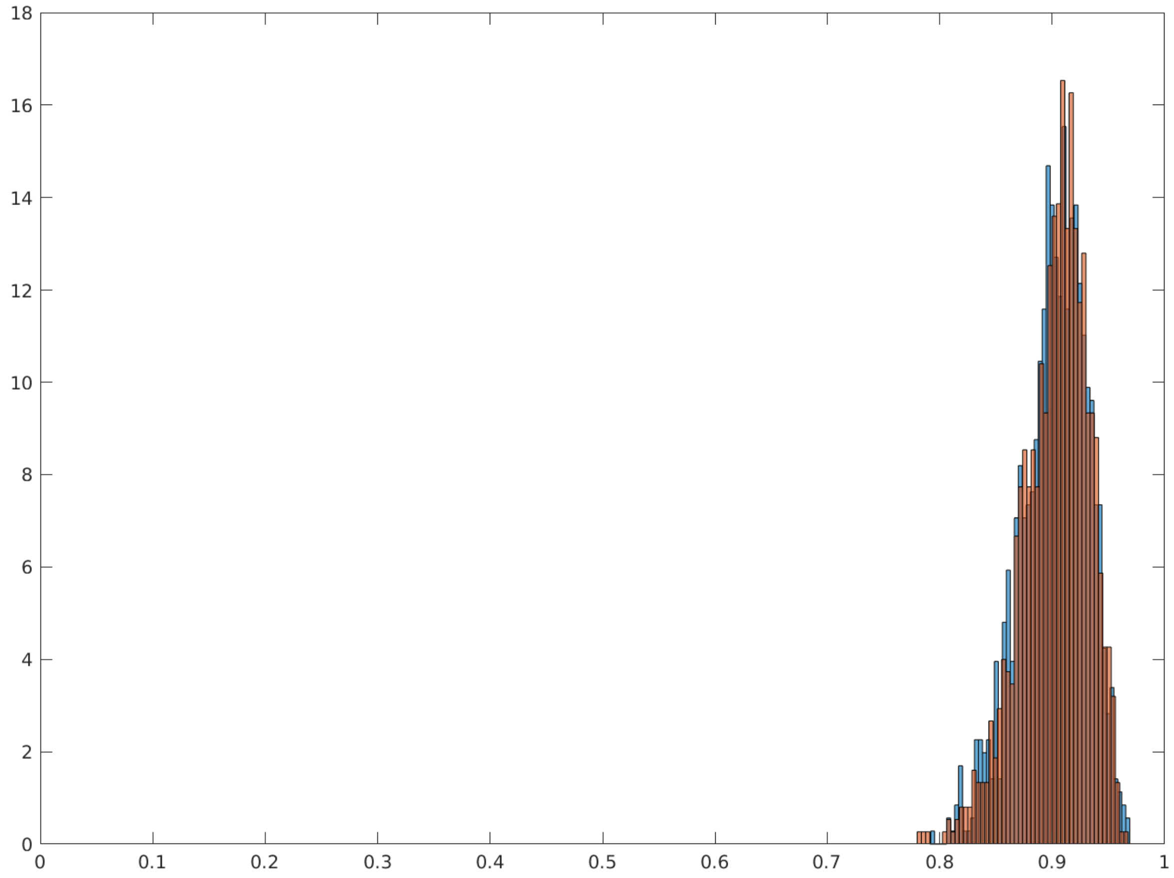
Markov-Chain Monte Carlo



https://chi-feng.github.io/mcmc-demo/

Example

For 100 data points:



Model Calibration

Experimental data:

$$\tilde{y}(x) = y(x) + e(x)$$

Measured response

True response

Measurement noise

Computational model:

$$\mathrm{y}(x) = f(x;\theta^*) + \delta(x)$$

Model

Model bias

Inputs:

$x \rightarrow system\ variables$

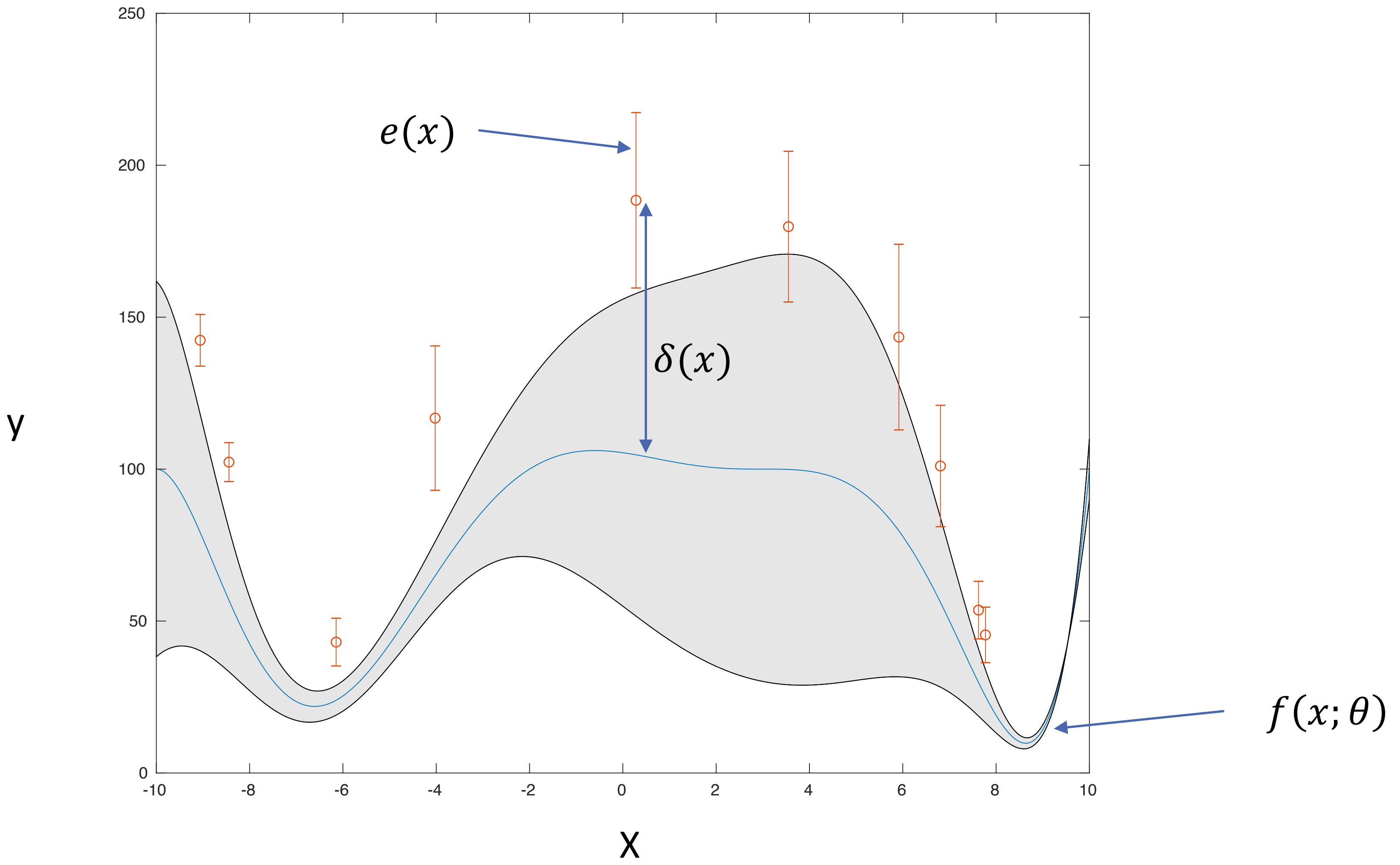
$\theta \rightarrow model\ parameters$

Errors:

$e(x) \rightarrow$ random variable of mean 0

$\delta(x) \rightarrow$ some unknown function

Model Calibration



Model Calibration

Experimental data:

$$\tilde{y}(x) = y(x) + e(x)$$

Measured response

True response

Measurement noise

Computational model:

$$\mathrm{y}(x) = f(x;\theta^*) + \delta(x)$$

Model

Model bias

Model updating equation:

$$\mathrm{y}(x) = f(x;\theta) + e(x)$$

Inputs:

$x \rightarrow system\ variables$

$\theta \rightarrow model\ parameters$

Errors:

$e(x) \rightarrow$ random variable of mean 0

$\delta(x) \rightarrow$ some unknown function

Transitional MCMC

Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging

Jianye Ching[1] and Yi-Chu Chen[2]

Abstract: This paper presents a newly developed simulation-based approach for Bayesian model updating, model class selection, and model averaging called the transitional Markov chain Monte Carlo (TMCMC) approach. The idea behind TMCMC is to avoid the problem of sampling from difficult target probability density functions (PDFs) but sampling from a series of intermediate PDFs that converge to the target PDF and are easier to sample. The TMCMC approach is motivated by the adaptive Metropolis–Hastings method developed by Beck and Au in 2002 and is based on Markov chain Monte Carlo. It is shown that TMCMC is able to draw samples from some difficult PDFs (e.g., multimodal PDFs, very peaked PDFs, and PDFs with flat manifold). The TMCMC approach can also estimate evidence of the chosen probabilistic model class conditioning on the measured data, a key component for Bayesian model class selection and model averaging. Three examples are used to demonstrate the effectiveness of the TMCMC approach in Bayesian model updating, model class selection, and model averaging.

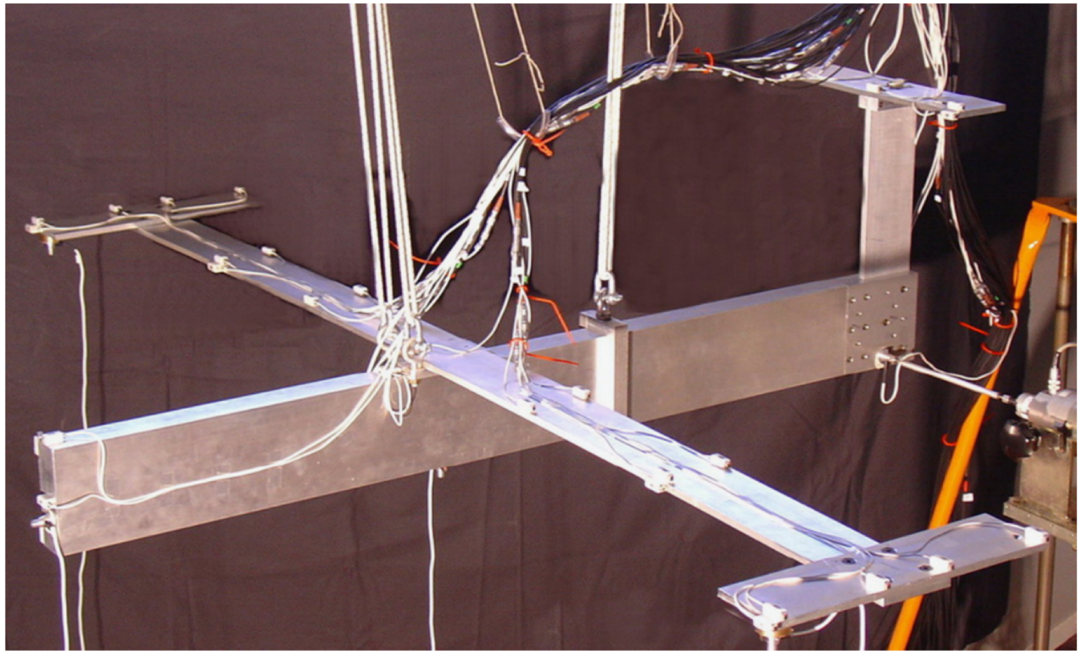
DOI: 10.1061/(ASCE)0733-9399(2007)133:7(816)

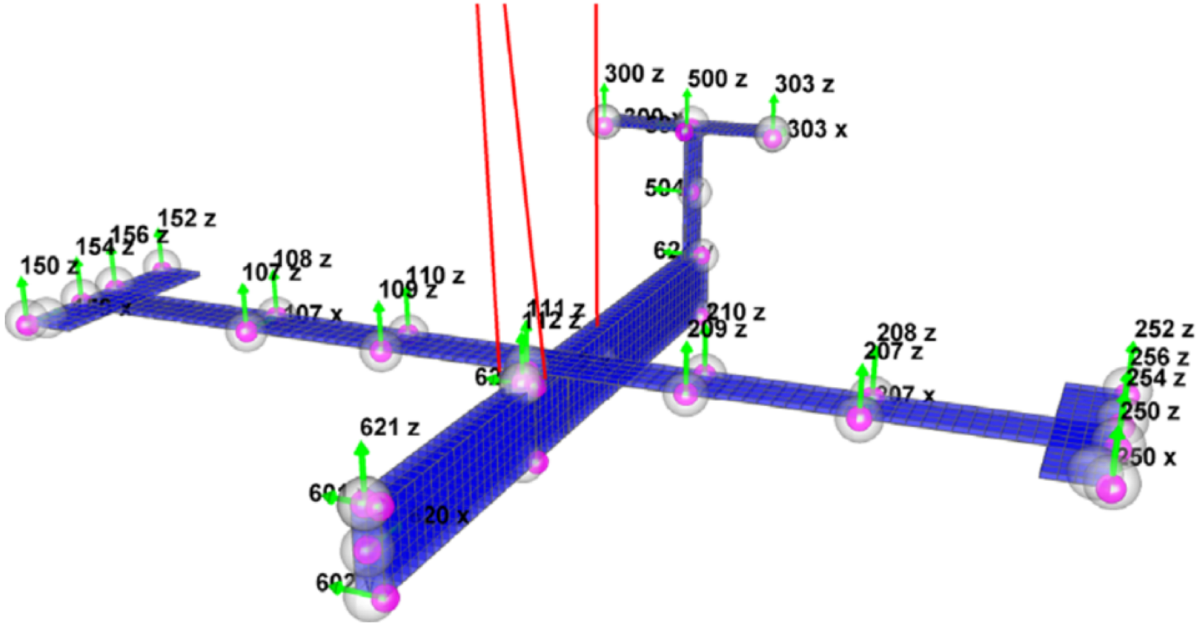
CE Database subject headings: Bayesian analysis; Simulation; Markov chains; Monte Carlo method.

Introduction

Model updating refers to the methodology that determines the most plausible model for an instrumented civil engineering system given its measured response and, possibly, its excitation. In recent years, civil engineers have paid much attention to model updating techniques since they have broad applications in structural health monitoring (Natke and Yao 1988; Hjelmstad and Shin 1997; Sanayei et al. 1999; Chang 2001; Beck et al. 2001; Casciati 2002; Bernal et al. 2002). Among the model updating techniques, when only some of the degrees of freedom (DOF) of the model are measured and when modeling errors are explicitly acknowledged. A previous Bayesian model updating approach (Beck and Katafygiotis 1998) has been successful in resolving the aforementioned difficulties when the updated (posterior) probability density function (PDF) of the model parameters is very peaked (e.g., the amount of data is sufficiently large) so that an asymptotic approximation of the Bayesian predictive integrals is accurate. However, when the PDF is not very peaked (e.g., limited data or a flat region in the PDF), the validity of the asymptotic approxi-

Example

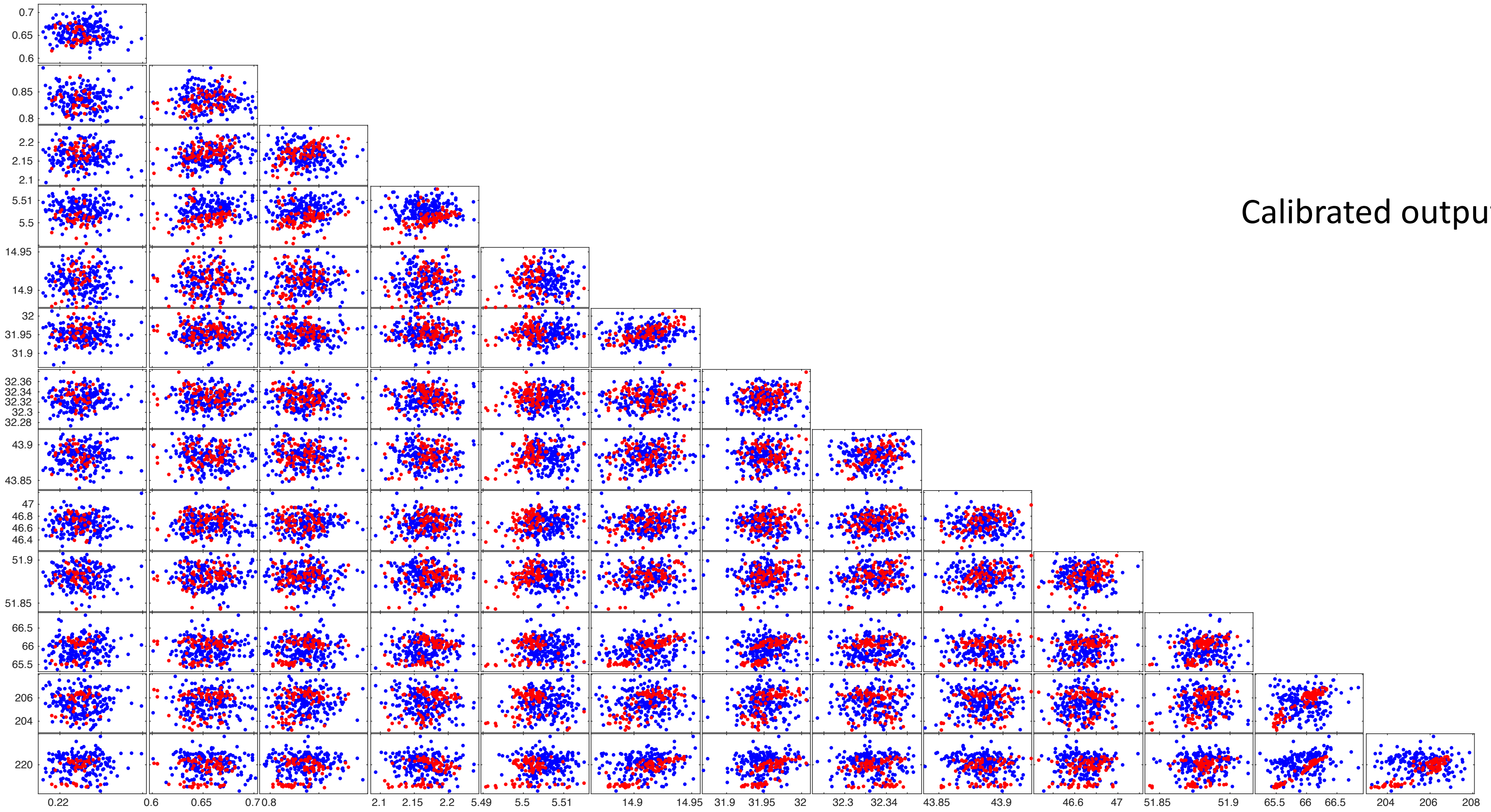




Input dimensions $\theta \rightarrow 18$

Output dimensions $y \rightarrow 14$

Calibrated outputs



Advantages and critiques

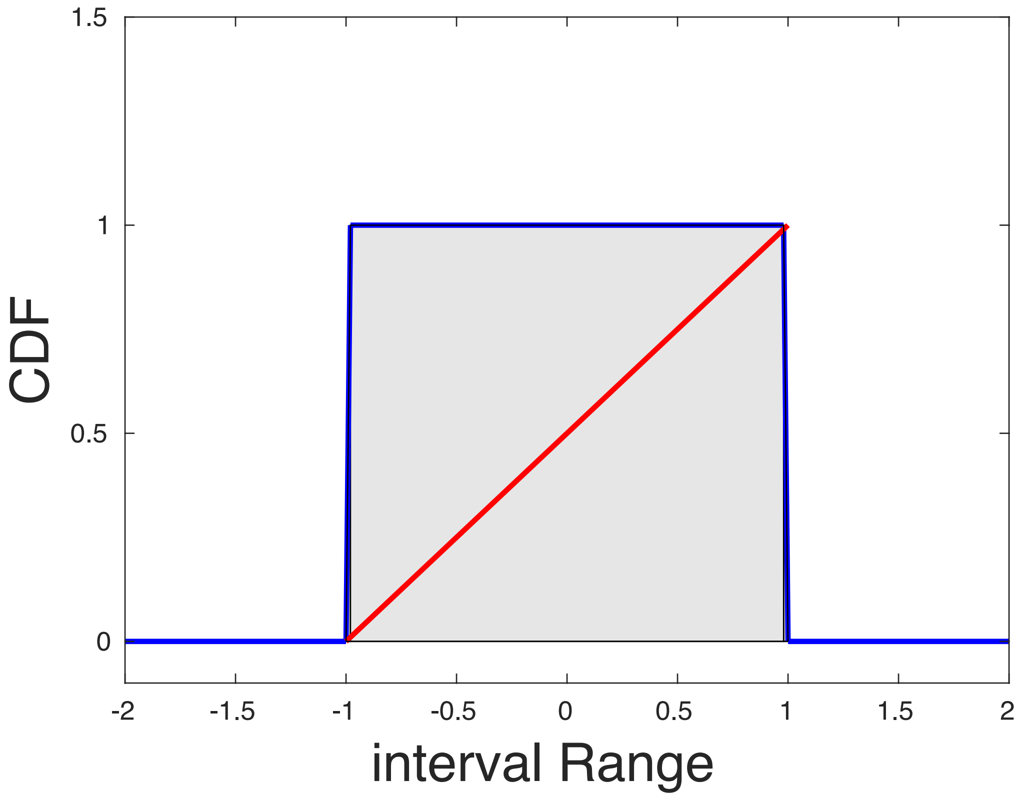
The good

- Calibrated model is uncertain, incorporating the calibration data uncertainty
- Inverse problem does not need to be globally identifiable
- Posteriors of other calculations can be reused or begin with expert opinion
- Easily parallelizable

The bad

- SUPER expensive, you may need emulators
- Implementation is tricky, can sometimes produce crap
- Likelihood and prior must be selected
- Uniform prior is not a state of maximum uncertainty

Advantages and critiques



The bad

- SUPER expensive, you may need emulators
- Implementation is tricky, can sometimes produce crap
- Likelihood and prior must be selected
- Uniform prior is not a state of maximum uncertainty

THE reference

J. R. Statist. Soc. B (2001)
63, *Part* 3, *pp.* 425–464

Bayesian calibration of computer models

Marc C. Kennedy and Anthony O'Hagan
University of Sheffield, UK

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, December 13th, 2000*, Professor P. J. Diggle *in the Chair*]

Summary. We consider prediction and uncertainty analysis for systems which are approximated using complex mathematical models. Such models, implemented as computer codes, are often generic in the sense that by a suitable choice of some of the model's input parameters the code can be used to predict the behaviour of the system in a variety of specific applications. However, in any specific application the values of necessary parameters may be unknown. In this case, physical observations of the system in the specific context are used to learn about the unknown parameters. The process of fitting the model to the observed data by adjusting the parameters is known as calibration. Calibration is typically effected by *ad hoc* fitting, and after calibration the model is used, with the fitted input values, to predict the future behaviour of the system. We present a Bayesian calibration technique which improves on this traditional approach in two respects. First, the predictions allow for all sources of uncertainty, including the remaining uncertainty over the fitted parameters. Second, they attempt to correct for any inadequacy of the model which is revealed by a discrepancy between the observed data and the model predictions from even the best-fitting parameter values. The method is illustrated by using data from a nuclear radiation release at Tomsk, and from a more complex simulated nuclear accident exercise.

Keywords: Calibration; Computer experiments; Deterministic models; Gaussian process; Interpolation; Model inadequacy; Sensitivity analysis; Uncertainty analysis

1. Overview

1.1. Computer models and calibration

Software for calibration and more

OpenCossan

University of Liverpool

DAKOTA

Explore and predict with confidence.

Sandia Labs, US