Applying Bayesian Statistical Methods to MICE

Robert D. Ryne

Center for Beam Physics
Accelerator Technology & Applied Physics Division
Lawrence Berkeley National Laboratory

June 22, 2015

Acknowledgements

- Chris Rogers (Rutherford Appleton Laboratory)
- Elizabeth Kelly, Earl Lawrence, David Higdon (Los Alamos National Laboratory)

What are Bayesian Methods?

Bayesian inference is a process of learning from data[1]

Bayesian statistical methods start with existing 'prior' beliefs, and update these using data to give 'posterior' beliefs, which may be used as the basis for inferential decisions[2]

[1]http://www.sagepub.com/upm-data/18550_Chapter6.pdf
[2]http://www.scholarpedia.org/article/Bayesian_statistics

Bayes Theorem

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

- $P(A|B)$: probability of A given B
- $P(B|A)$: probability of B given A
- $P(A)$: prob of A
- $P(B)$: prob of B

As used in Bayesian Statistics:

$$\pi(\theta|y) \propto L(y|\eta(x,\theta)) \times \pi(\theta)$$

posterior likelihood prior

The posterior is proportional to the likelihood times the prior

likelihood: probability of observing output y given a model η that depends on observable inputs x and parameters θ

Bayesians vs Frequentists*

Where appreciable prior information exists, perhaps the most significant difference between Bayesian and frequentist methods is the ability of the Bayesian analysis to make use of that additional info [the prior dist]

As a result, Bayesian methods will typically produce stronger inferences from the same data

Furthermore, the prior information allows the Bayesian analysis to be more responsive to the context of the data

However, the prior distribution is also the focus of opposition to Bayesian methods from adherents of the frequentist philosophy

Frequentists regard its use as unscientific, so do not believe that such stronger or more responsive inferences can be obtained legitimately

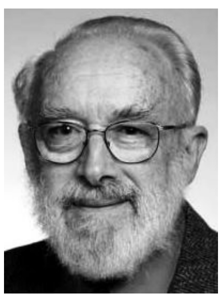
*http://www.sagepub.com/upm-data/18550_Chapter6.pdf

Bayesians vs Frequentists*

Two old-timers slugging out the Bayes *vs* Frequentist battle;

If [Bayesians] would only do as [Bayes] did and publish posthumously we should all be saved a lot of trouble

Maurice Kendall (1907–1983), JRSSA 1968



The only good statistics is Bayesian Statistics

Dennis Lindley (1923–2013)
in 'The Future of Statistics: A Bayesian 21st Century' (1975)

*http://faculty.washington.edu/kenrice/BayesIntroClassEpi515.pdf

Bayesians vs Frequentists

A Bayesian is

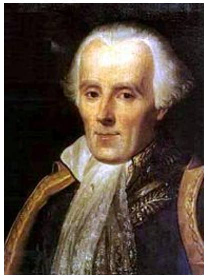
*One who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes he has seen a mule**

*Stephen Senn, Statistical Issues in Drug Development, 2nd Ed., Wiley & Sons Ltd., © 2007

Why now?

- Roots of Bayesian methods ~1700's
- Growth in late 1980's / 1990's

Bayes



Laplace

*In their highly influential JASA 1990 paper, Alan Gelfand and Adrian Smith projected the Bayesian paradigm towards the stars when they recommended Markov Chain Monte Carlo (MCMC) simulations as a way of computing Bayesian estimates and inferences for the parameters in a wide range of complicated sampling models, in situations where it was well-nigh impossible to achieve a solution using ordinary Monte Carlo or Importance Sampling techniques**

*http://www.thomashoskynsleonard.co.uk/personal_history_6.htm

What does this have to do with MICE?

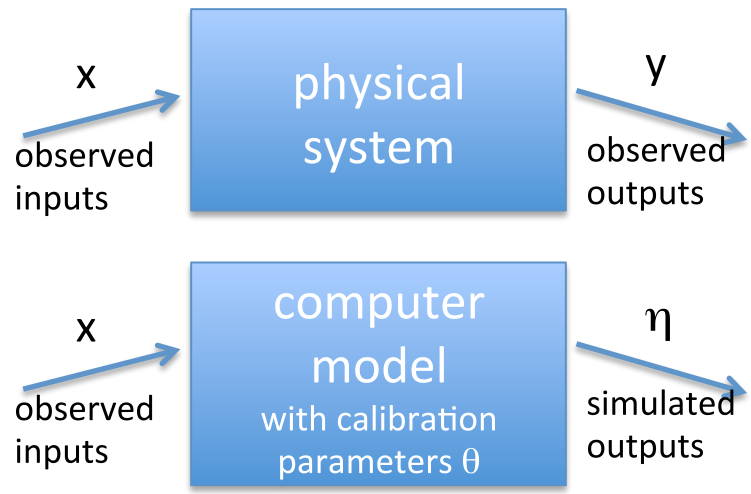
Prior belief about the beamline and instruments

Computer model (MAUS)

Data

Computer model calibration, inference, prediction

Bayesian Inference for Code Calibration*



$y(x_i)=\eta(x_i,\theta)+\varepsilon(x_i)+\delta(x_i)$ $i=1,\ldots,n$ i denotes a measurement

- At various inputs x, we have measurements y, with measurement error ε and model error δ
- We have a sampling model or likelihood, L, for y

*D. Higdon et al., "Combining Field Data and Computer Simulations for Calibration and Prediction," SIAM J. Sci. Comput. Vol. 26, No. 2, pp. 448-466 (2004)

Bayesian Inference, cont.

- Suppose we have some prior knowledge about what we think the model parameters, θ, must be for the simulator to agree with measurements. Let $\pi(\theta)$ denote the prior distribution.
- The Bayesian formulation states:

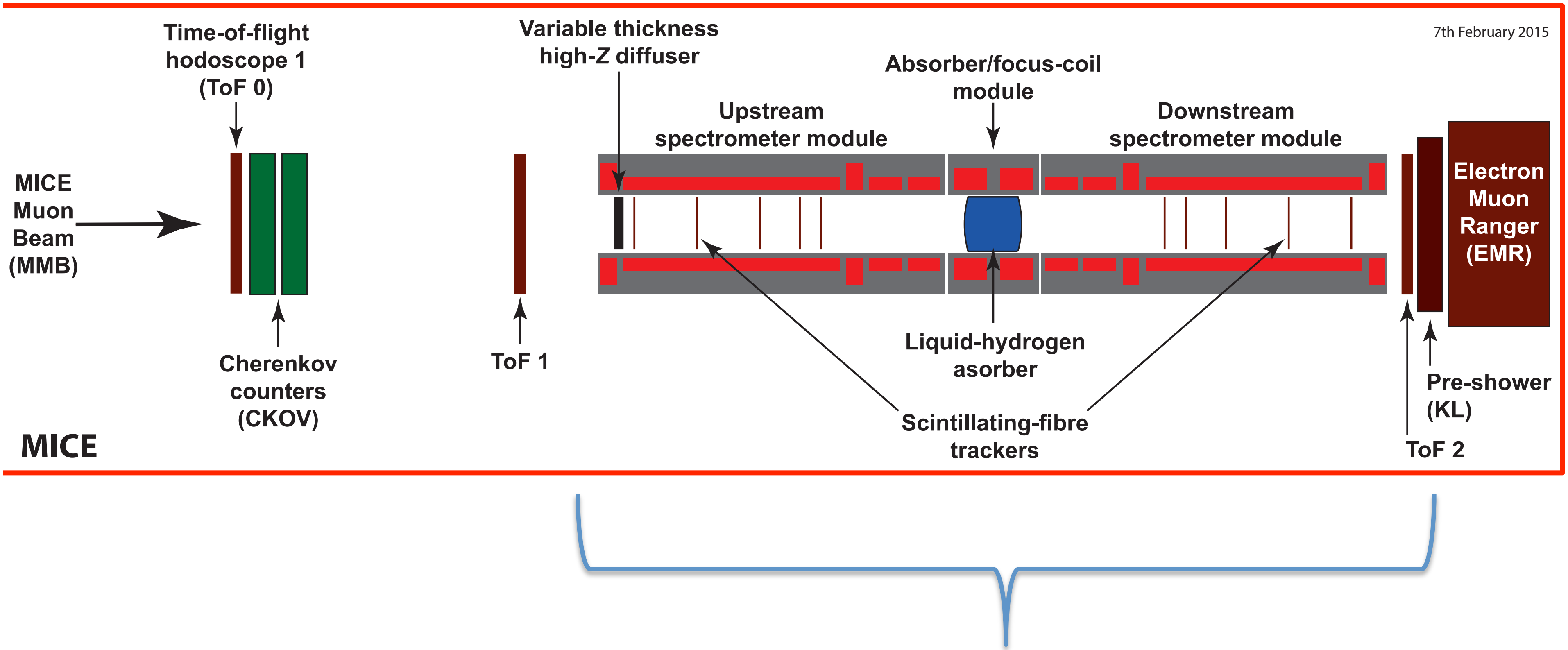
$$\pi(\theta|y) \propto L(y|\eta(x,\theta)) \times \pi(\theta)$$

The posterior is proportional to the likelihood times the prior

- Instead of thinking of θ simply as an ordinary scalar or vector quantity, we *think of θ as a random variable with a distribution associated with it*
  - initially, this is the prior $\pi(\theta)$
- We want to find the posterior distribution of θ given the data y, $\pi(\theta|y)$
  - Then we can determine moments of $\pi(\theta|y)$, Bayesian credible intervals, etc.
- Also can determine posterior for variance parameters of sampling model, $\pi(\Sigma|y)$

All this involves combining observations with computer simulations

MICE Step IV



We want to calibrate a computer model of this

Inference of 10 solenoid parameters and 4 measurement uncertainties

	parameter	exact	prior mean	posterior mean	prior std. dev.	posterior std. dev.
current density	θ_1	151.634	147.	151.623	40.	.0185
	θ_2	123.807	131.	123.752	40.	.0615
	θ_3	142.602	135.	142.762	40.	.0722
	θ_4	118.863	113.	118.930	40.	.0496
	θ_5	103.874	104.	103.743	40.	.0652
	θ_6	-101.920	-104.	-101.668	40.	.0918
	θ_7	-108.330	-112.	-108.203	40.	.0753
	θ_8	-132.950	-140.	-132.786	40.	.0976
	θ_9	-127.378	-131.	-127.736	40.	.1266
	θ_{10}	-133.948	-147.	-134.162	40.	.0669
1/variance	τ_1	6.250e6	5.e6	6.256e6	1.0e6	.0903e6
	τ_2	2500.	5000.	2434.	2236.	33.8
	τ_3	6.250e6	5.e6	6.351e6	1.0e6	.0867
	τ_4	2500.	5000.	2508.	2236.	36.7

# of observations=10000 length of MCMC chain = 5000 after 5000 burn-in

Bayesian inference of the 4x4 Linear Map: Results

MaryLie/Impact ("true") 4x4 matrix:

$$\begin{bmatrix} 4.81344E-01 & -1.55775E-01 & -9.60235E-03 & 3.10757E-03 \\ 4.83031E+00 & 5.13476E-01 & -9.63599E-02 & -1.02433E-02 \\ 9.60235E-03 & -3.10757E-03 & 4.81344E-01 & -1.55775E-01 \\ 9.63599E-02 & 1.02433E-02 & 4.83031E+00 & 5.13476E-01 \end{bmatrix}$$

MCMC results (10K observations; MCMC length=75K+75K burn-in:

$$\begin{bmatrix} 4.81349E-01 & -1.55757E-01 & -9.60723E-03 & 3.11229E-03 \\ 4.82935E+00 & 5.13961E-01 & -9.65443E-02 & -1.03595E-02 \\ 9.65088E-03 & -3.11962E-03 & 4.81362E-01 & -1.55776E-01 \\ 9.66604E-02 & 1.02018E-03 & 4.82893E+00 & 5.13895E-01 \end{bmatrix}$$

Both of these matrices are symplectic

	μ_{post}	σ_{post}
θ_2	1.556	2.04e-4
θ_5	-0.299	2.86e-5
θ_7	-1.374	1.00e-4
θ_9	-2.000e-2	3.83e-5

Turns out (due to symmetry of MICE channel) that only 4 regression coefficients matter.
Note the small σ compared to the mean of these 4.

	μ_{post}	σ_{post}
θ_1	-9.e-6	2.8e-5
θ_3	-1.5e-5	2.9e-5
θ_4	4.6e-5	1.5e-4
θ_6	-8.8e-6	2.0e-4
θ_8	-1.3e-5	9.8e-5
θ_{10}	-2.6e-6	7.1e-5

Conclusions

- Bayesian techniques are extremely powerful and flexible
- Applied to MICE, they can be used for computer model calibration, to infer the transfer map, to predict the impact of changes, to test ideas, and to provide insight
- The examples here demonstrate how measurements and simulation can be combined to
  - infer model parameters, (e.g. magnet current settings) so that the computer model agrees with expt, *including distributions that describe the uncertainty of inferred parameters*
  - infer the measurement uncertainty
  - infer the transfer map
- The techniques should be broadly applicable to other accelerator experiments as well

On-line example code

To try this yourself, you can download a sample code from

`http://portal.nersc.gov/project/m669/bayes5term.f90`

Simulator: y=p1 + p2 x + p3 x^2 + p4 x^3 + p5 x^4 +ε   (p1-p5 are calibration params)

7500 observations, observation error σ = 0.25; 40000 MCMC steps