An Example of Bayesian Model Calibration

Earl Lawrence

LANL

Motivation

Is it possible to make scientific inference with a model when the parameters are unknown and the model might be wrong and the data is terrible?

Thomas claims:

1. The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2d on supposition the 1st happens.
2. If there be two subsequent events be determined every day, and each day the probability of the 2d is b/N and the probability of both P/N, and I am to receive N if both of the events happen the 1st day on which the 2d does; I say, according to these conditions, the probability of my obtaining N is P/b.

Bayes, T. and Price, R. (1763) "An Essay towards solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society of London.* 53:370-418.

Los Alamos NATIONAL LABORATORY EST. 1943

Thomas claims:

1. $P(A, B) = P(A)P(B|A)$
2. $P(A|B) = P(A, B)/P(B)$

$$P(A|B) = P(A)P(B|A)/P(B)$$

Thomas claims:

$$p(\theta|y) \propto \pi(\theta) f(y|\theta)$$

- θ is a parameter vector, or a model, or something we want to learn
- y is data
- $p(\cdot)$ is the posterior
- $\pi(\cdot)$ is the prior
- $f(\cdot)$ is the likelihood

Simple Gaussian Example

- $y = \theta + \epsilon, \;\; \epsilon \sim N(0, \sigma^2)$
- $\theta \sim N(\mu, \delta^2)$
- Observe $y_1, \cdots, y_n$

$$
\begin{aligned}
p(\theta|y) &\propto \exp\left\{-\frac{1}{2\delta^2}(\theta-\mu)^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\sum_i (y_i-\theta)^2\right\} \\
&\;\;\vdots \\
\theta|y &\sim N\left(\nu, \gamma^2\right) \\
\nu &= \frac{\frac{n}{\sigma^2}\bar{Y} + \frac{1}{\delta^2}\mu}{\frac{n}{\sigma^2} + \frac{1}{\delta^2}} \\
\gamma^2 &= \left(\frac{n}{\sigma^2} + \frac{1}{\delta^2}\right)^{-1}
\end{aligned}
$$

Los Alamos NATIONAL LABORATORY EST. 1943

Markov Chain Monte Carlo

Method of drawing sequence of correlated samples from a distribution by constructing a Markov chain whose stationary distribution is the one of interest. This is useful when the distribution is not otherwise tractable and only requires knowing the distribution up to a constant. The samples can be used for inference (e.g. means, variances quantiles).

Markov Chain Monte Carlo: Metropolis-Hastings

Assume x follows some distribution with density p and that we have x_k with $p(x_k) > 0$.

1. Draw a candidate x' from $q(x'|x_k)$.
2. Compute $\alpha = \frac{p(x')q(x_k|x')}{p(x_k)q(x'|x_k)}$
3. Draw $u \sim Unif(0,1)$.
4. If $u \leq \alpha$, set $x_{k+1} = x'$, else set $x_{k+1} = x_k$.

Often, q is a random walk so $q(x'|x_k) = q(x_k|x')$ and α simplifies (original Metropolis). Sometimes, q is p (Gibbs sampling). Good results often require some tuning of q (e.g. the step size of the random walk).

Los Alamos NATIONAL LABORATORY EST. 1943

Simple Gaussian Example with MCMC

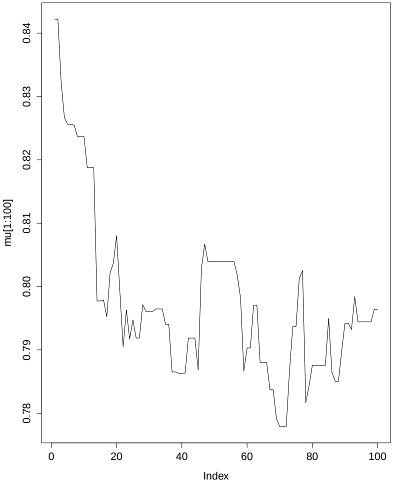


Figure: First 100 draws μ.

Simple Gaussian Example with MCMC

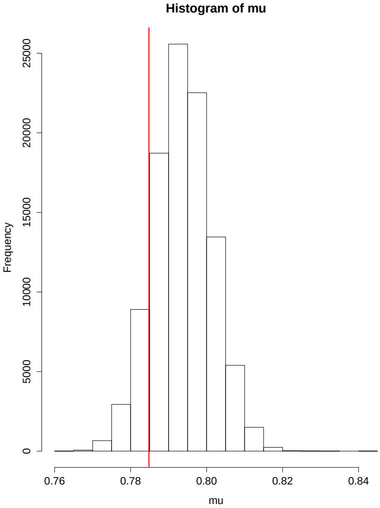


Figure: Histogram of μ with "true" value.

Simple Gaussian Example with MCMC

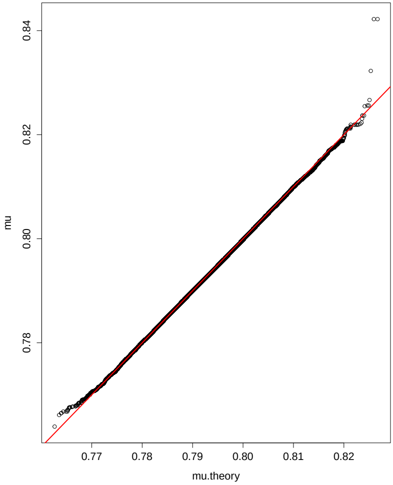


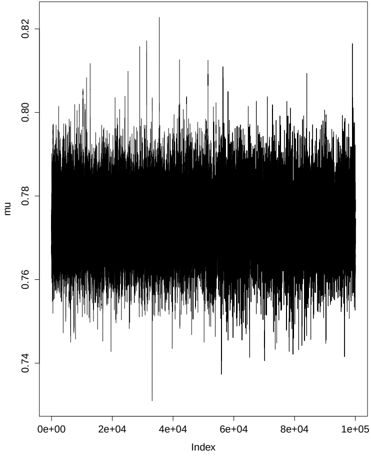
Figure: Comparison of the MCMC result and they theoretical result.

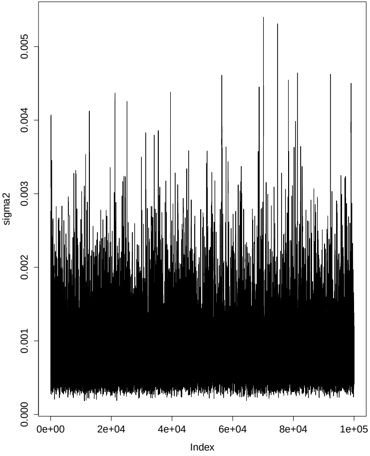
Gaussian Example with Unknown Mean and Variance

- $y|\theta \sim N(\theta, \sigma^2)$
- $\theta \sim N(\mu, \delta^2)$
- $\sigma^2 \sim Unif(0, U)$
- Observe $y_1, \cdots, y_n$

Sample from $p(\mu, \sigma^2|y)$ by sampling sequentially from the full conditional posteriors: $p(\mu|\sigma, y)$ and $p(\sigma^2|\mu, y)$ which are simply proportional to their joint density.

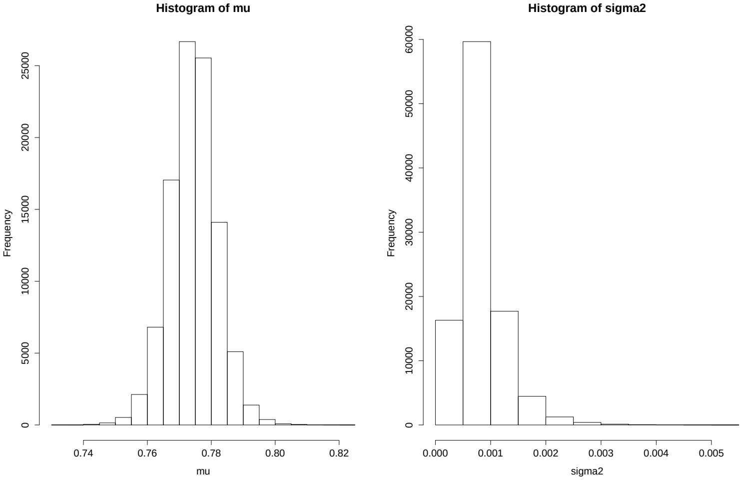
Gaussian Example with Unknown Mean and Variance





Los Alamos
NATIONAL LABORATORY
EST. 1943

Gaussian Example with Unknown Mean and Variance

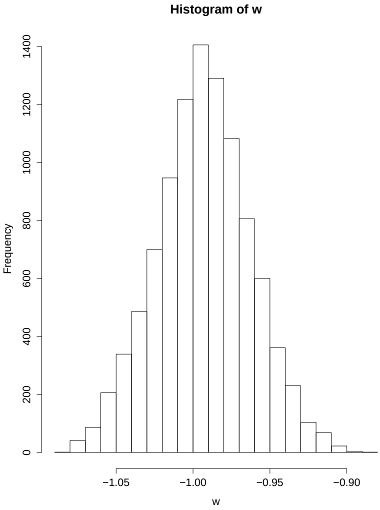


Black Box Functions

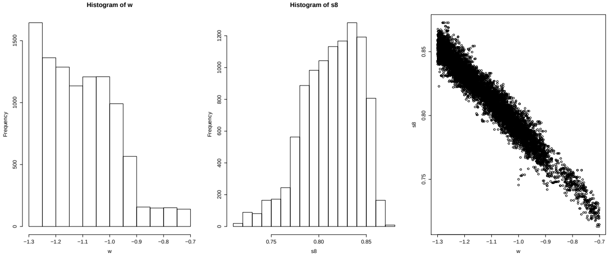
- $y|\theta \sim N(\eta(\theta), \sigma^2)$
- $\theta \sim \pi(\theta)$

MCMC only requires that you can evaluate $\eta(\cdot)$.

Black Box 1-D: Mini Cosmic Emu with Unknown w



Black Box 2-D: Mini Cosmic Emu with Unknown w and σ_8



Los Alamos NATIONAL LABORATORY EST.1943

Slow Black Box Functions

What if the function takes a month of computation? When the simulation is to slow to call inside the MCMC, we need to approximate it. The basic idea is to run the simulation over a training set and build a statistical model that predicts the results at untried settings.

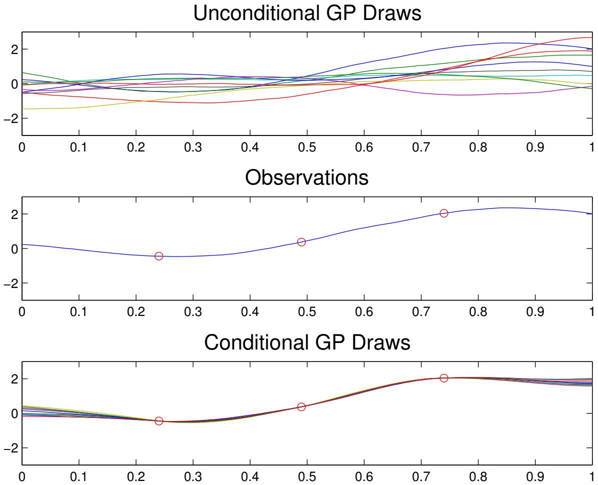
Gaussian Process

Assume that univariate y is a function of d-D x. Let $\vec{y}$ be a collection of these points associated with the matrix X (ith row goes with y_i).

$$\begin{aligned}\vec{y} &\sim N\left(\vec{0}, \sigma^2 R(X)\right) \\ R_{i,j} &= \exp\left\{-\sum_{k=1}^{p} \beta_k (X_{i,k} - X_{j,k})^2\right\}\end{aligned}$$

This has the squared exponential covariance which produces continuous and very smooth draws. Given a training set $(\vec{y}, X)$ and priors, the GP parameters σ^2 and $\vec{\beta}$ can be estimated with MCMC.

Gaussian Process Cartoon



Conditional GP

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{\vec{0}, \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right]\right\} \quad (1)$$

$$y_1|y_2 \sim N\left\{\Sigma_{12}\Sigma_{22}^{-1}y_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right\} \quad (2)$$

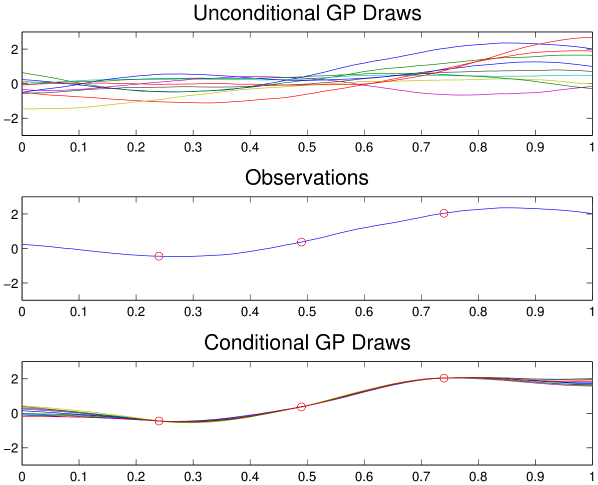
Assume that Σ is the aforementioned function of X, that y_2 are points that we've observed at X_2, and that y_1 are points that we want to predict at X_1. Everything on the right is known and gives us the distribution for the new points.

Conditional GP

$$y_1 | y_2 \sim N \left\{ \Sigma_{12} \Sigma_{22}^{-1} y_2, \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right\}$$

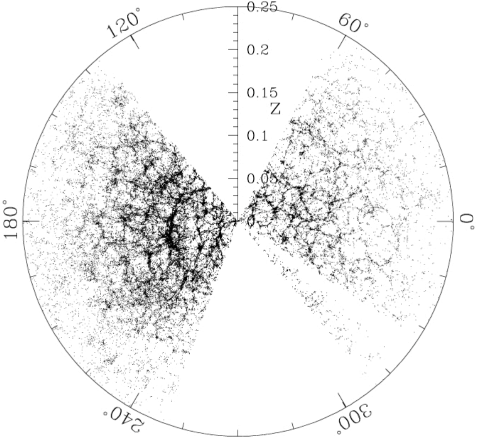
- The mean for new points is a weighted average of the observed points.
- The variance goes to zero as a new point approaches an observed point.
- This is just Bayes rule again. The GP is a prior for the unobserved points and we know the conditional relationship between the observed points and unknown points.

Gaussian Process Cartoon



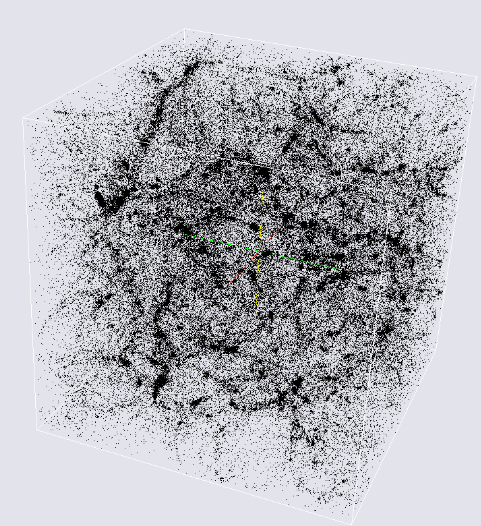
The Universe. How's that work?

Sloan Digital Sky Survey.



Simulating the Universe

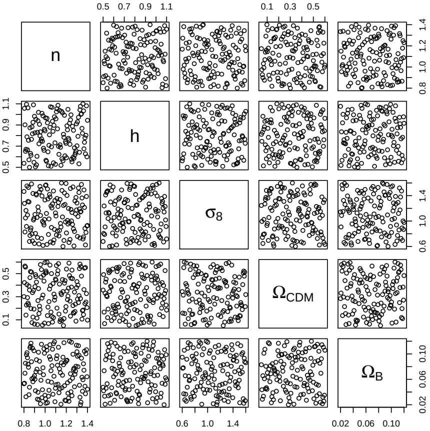
N-Body Simulations. These take a long time.



Los Alamos NATIONAL LABORATORY EST. 1943

Choosing the Simulations

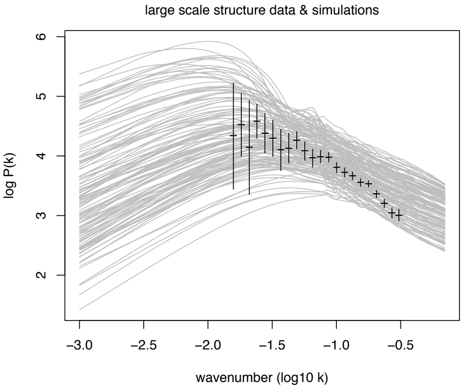
Latin Hypercube over the parameters of interest.



Los Alamos
NATIONAL LABORATORY
EST. 1943

Comparing Sims and Data

The power spectrum describes how the matter is distributed over large scales.



Los Alamos NATIONAL LABORATORY EST. 1943

Making the Sausage, Step 1

The data is a noisy version of the model at the "correct" input.

$$
\begin{aligned}
y &= \eta(\theta) + \epsilon \\
\epsilon &\sim \mathrm{N}\left(\vec{0}, \Sigma_y\right) \\
\pi(\theta) &= 1, \theta \in \mathcal{C} \\
p(\theta|y) &\propto \exp\left\{\frac{1}{2}(y - \eta(\theta))'\Sigma_y^{-1}(y - \eta(\theta))\right\}
\end{aligned}
$$

Making the Sausage, Step 2

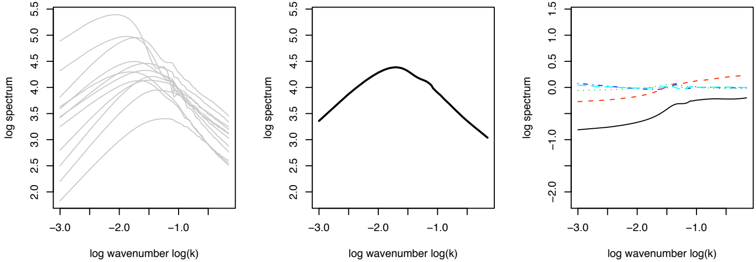
Treat $\eta(\cdot)$ as an unknown function, with observations $\eta^* = (\eta(t_1), \ldots, \eta(t_m))'$.

$$\pi(\theta, \eta(\cdot)|y, \eta^*) \propto L(y|\eta(\theta)) \cdot L(\eta^*|\eta(\cdot)) \cdot \pi(\eta(\cdot)) \cdot \pi(\theta)$$

η: Decomposing the Multivariate Output

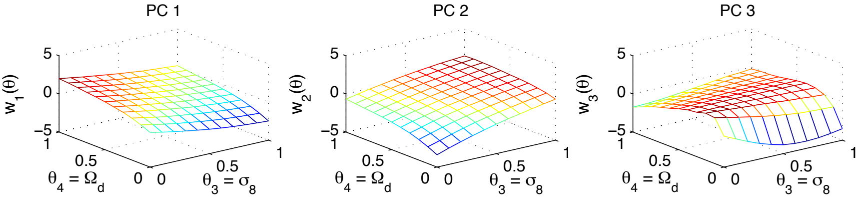
Compute a principal component basis from the simulations.

$$\eta(t) = \sum_{i=1}^{q} \phi_i w_i(t) + \epsilon$$



η: GPs for Basis Weights

$$
\begin{aligned}
w_i(t) &\sim \mathrm{N}(0, \lambda_{wi}^{-1} R(t; \rho_i)) \\
\mathrm{Corr}(w_i(t), w_i(t')) &= \prod_{k=1}^{p} \rho_{ik}^{4(t_k - t'_k)^2}
\end{aligned}
$$



Los Alamos
NATIONAL LABORATORY
EST. 1943

η Start Combining Things

$$\begin{pmatrix} w_1 \\ \vdots \\ w_q \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda_{w1}^{-1} R(t;\rho_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{wq}^{-1} R(t;\rho_q) \end{pmatrix} \right)$$

Find η again

$$\eta | w, \lambda_\eta \sim N\left(\Phi w, \frac{1}{\lambda_\eta} I \right)$$

η Put some priors on things

$$
\begin{aligned}
\pi(\lambda_{wi}) &\propto \lambda_{wi}^{a_w-1} e^{-b_w \lambda_{wi}}, \quad i = 1, \ldots, q, \\
\pi(\rho_{ik}) &\propto \rho_{ik}^{a_\rho - 1}(1-\rho_{ik})^{b_\rho - 1}, \quad i = 1, \ldots, q,\ k = 1, \ldots, p
\end{aligned}
$$

Los Alamos NATIONAL LABORATORY EST. 1943

Making the Sausage, Step 3: Just the emulator.

$$\pi(\lambda_\eta, \lambda_w, \rho | \eta) \propto$$

$$\left|(\lambda_\eta \Phi'\Phi)^{-1} + \Sigma_w\right|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\hat{w}'([\lambda_\eta \Phi'\Phi]^{-1} + \Sigma_w)^{-1}\hat{w}\} \times$$

$$\lambda_\eta^{a_\eta^*-1} e^{-b_\eta^* \lambda_\eta} \times \prod_{i=1}^{q} \lambda_{wi}^{a_w-1} e^{-b_w \lambda_{wi}} \times \prod_{i=1}^{q}\prod_{j=1}^{p} \rho_{ij}^{a_\rho-1}(1-\rho_{ij})^{b_\rho-1},$$

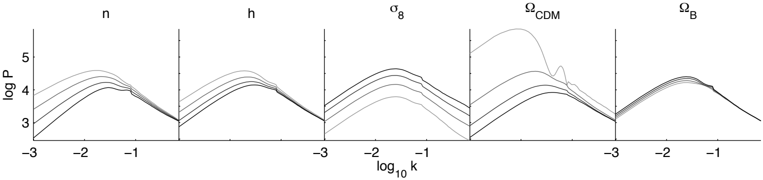
where

$$\begin{aligned} a_\eta^* &= a_\eta + \frac{m(n_\eta - q)}{2}, \\ b_\eta^* &= b_\eta + \tfrac{1}{2}\eta'(I - \Phi(\Phi'\Phi)^{-1}\Phi')\eta, \text{ and} \\ \hat{w} &= (\Phi'\Phi)^{-1}\Phi'\eta. \end{aligned}$$

Los Alamos NATIONAL LABORATORY EST. 1943

The emulator is pretty useful.

For example, it can do sensitivity studies.



Bringing data back

$$\begin{aligned} y &= \eta(\theta) + \epsilon, \\ y &= \Phi_y w(\theta) + \epsilon \\ y|w(\theta), \lambda_y &\sim N(\Phi_y w(\theta), (\lambda_y W_y)^{-1}),\ \lambda_y \sim Ga(a_y, b_y) \end{aligned}$$

Making the Sausage, Step 4 (Part 1)

$$
\begin{aligned}
\hat{w}_y &= (\Phi_y' W_y \Phi_y)^{-1} \Phi_y' W_y y, \\
a_y^* &= a_y + \tfrac{1}{2}(n - q), \\
b_y^* &= b_y + \tfrac{1}{2}(y - \Phi_y \hat{w}_y)' W_y (y - \Phi_y \hat{w}_y), \\
\Lambda_y &= \lambda_y \Phi_y' W_y \Phi_y, \\
\Lambda_\eta &= \lambda_\eta \Phi' \Phi, \\
I_q &= q \times q \text{ identity matrix}, \\
\Sigma_{w_y w} &= \begin{pmatrix} \lambda_{w1}^{-1} R(\theta, \theta^*; \rho_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{wq}^{-1} R(\theta, \theta^*; \rho_q) \end{pmatrix}, \\
\hat{z} &= \begin{pmatrix} \hat{w}_y \\ \hat{w} \end{pmatrix}, \\
\Sigma_{\hat{z}} &= \begin{pmatrix} \Lambda_y^{-1} & 0 \\ 0 & \Lambda_\eta^{-1} \end{pmatrix} + \begin{pmatrix} I_q & \Sigma_{w_y w} \\ \Sigma_{w_y w}' & \Sigma_w \end{pmatrix}.
\end{aligned}
$$

Making the Sausage, Step 4 (Part 2)

$$\pi(\lambda_\eta, \lambda_w, \rho, \lambda_y, \theta | \hat{z}) \propto$$

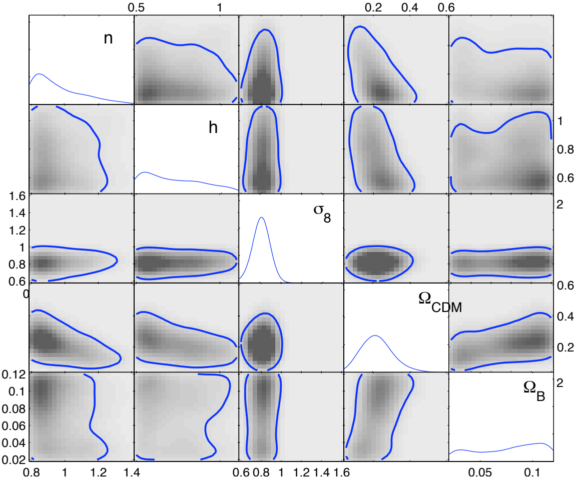
$$|\Sigma_{\hat{z}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\hat{z}'\Sigma_{\hat{z}}^{-1}\hat{z}\right\} \times \lambda_\eta^{a_\eta^*-1} e^{-b_\eta^* \lambda_\eta} \times \prod_{i=1}^{q} \lambda_{wi}^{a_w-1} e^{-b_w \lambda_{wi}} \times$$

$$\prod_{i=1}^{q}\prod_{k=1}^{p} \rho_{ik}^{a_\rho-1}(1-\rho_{ik})^{b_\rho-1} \times \lambda_y^{a_y^*-1} e^{-b_y^*\lambda_y} \times I[\theta \in C],$$

Cooking the Sausage

All of these parameters are estimated with one-at-a-time Metropolis-Hastings MCMC.

Posterior for Scientific Parameters



Higdon, Lawrence, Heitmann, Habib (2012).

Los Alamos NATIONAL LABORATORY EST. 1943

Stuff Not Appearing In This Talk

- Prediction, with all of the uncertainty.
- Systematic bias.