LA-UR-15-25984
Approved for public release; distribution is unlimited.

Title: Basics of Bayesian Statistics and Emulation

Author(s): Lawrence, Earl Christopher

Intended for: Santa Fe Cosmology Workshop, 2015-07-24 (Santa Fe, New Mexico, United States)

Issued: 2015-07-29

Basics of Bayesian Statistics and Emulation

Earl Lawrence

LANL

Motivation

How can we make statistical inferences when our predictions are given by computationally intensive simulations?

Bayesian Inference

- The goal is to learn about unknowns θ from observables y.
- We have some idea about the unknowns captured in our prior distributions $\pi(\theta)$.
- We also have some idea about how the observables depend on the unknowns given by our likelihood $f(y|\theta)$.
- Bayes Theorem: $p(\theta|y) \propto \pi(\theta) f(y|\theta)$.

Simple Gaussian Example

- $y|\theta \sim N(\theta, \sigma^2)$
- $\theta \sim N(\mu, \delta^2)$
- 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 $u \sim \mathit{Unif}(0, 1)$.
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).

Simple Gaussian Example with MCMC

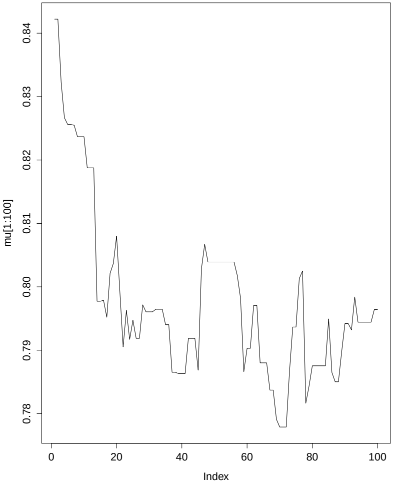


Figure: First 100 draws μ.

Los Alamos NATIONAL LABORATORY EST. 1943

Simple Gaussian Example with MCMC

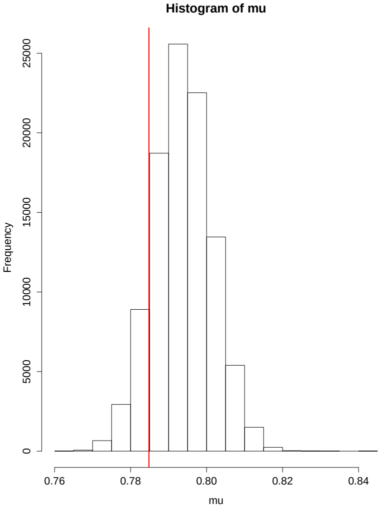


Figure: Histogram of μ with "true" value.

Simple Gaussian Example with MCMC

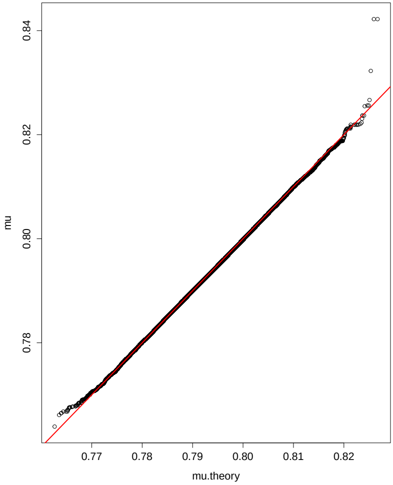


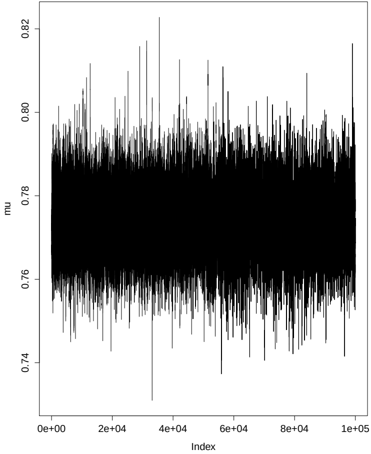
Figure: Comparison of the MCMC result and they theoretical result.

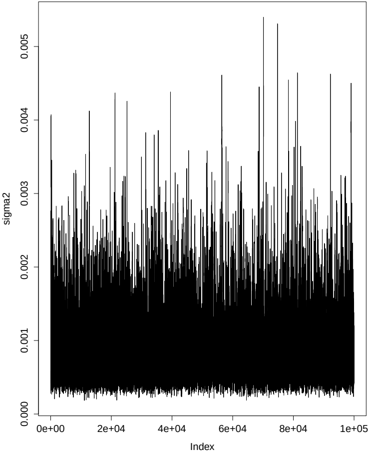
Gaussian Example with Unknown Mean and Variance

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

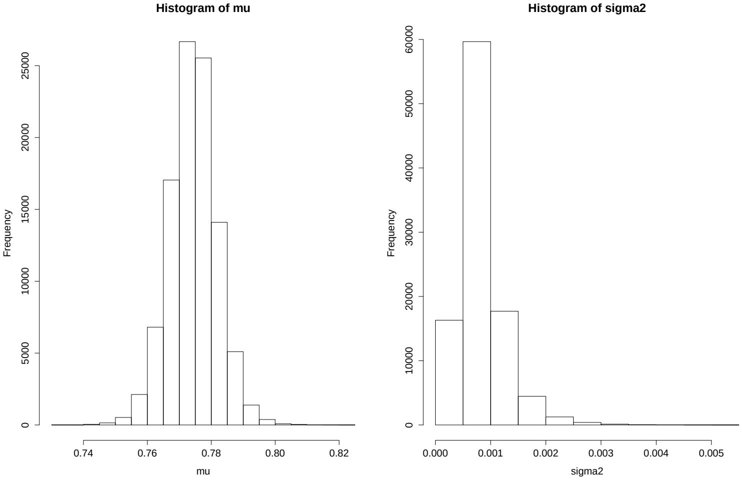
Sample from $p(\mu, \sigma^2|y)$ by sampling sequentially from the full conditional posteriors: $p(\mu|\sigma, y)$ and $p(\sigma^2|\mu, y)$ which are simply proportional to their joint density.

Gaussian Example with Unknown Mean and Variance





Gaussian Example with Unknown Mean and Variance

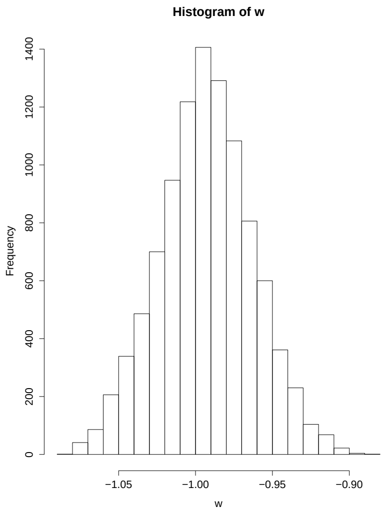


Black Box Functions

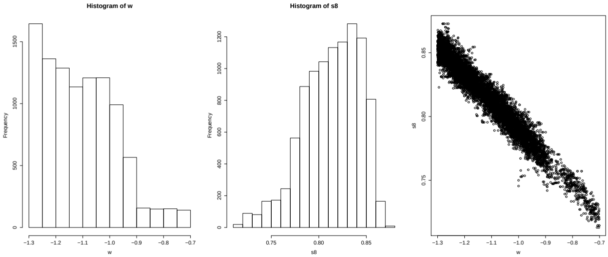
- $y|\theta \sim N(\eta(\theta), \sigma^2)$
- $\theta \sim \pi(\theta)$

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



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 interpolates 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.

Conditional GP

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

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

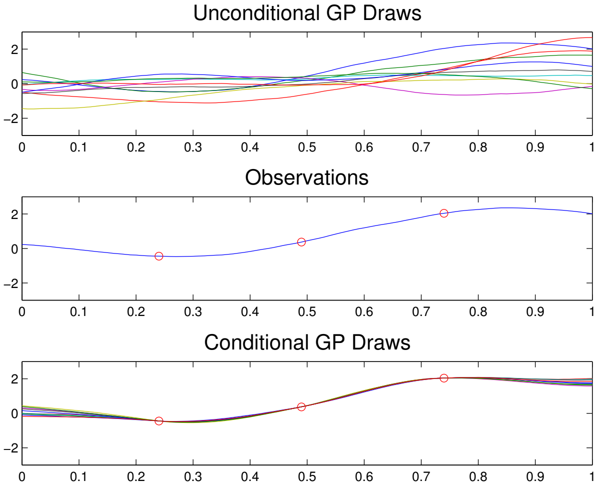
Assume that Σ is the aforementioned function of X, that y_2 are points that we've observed at X_2, and that y_1 are points that we want to predict at X_1. Everything on the right is known and gives us the distribution for the new points.

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 dot product.
- 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



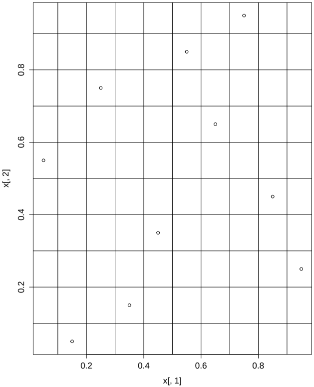
Los Alamos
NATIONAL LABORATORY
EST. 1943

Design for GP

The variance at a unobserved point depends on the distance to the training points. Thus, we want the training points to be space-filling. There are numerous variations.

Design for GP: Latin Hypercubes

For n points in p dimensions, divide each dimension into n bins, get p permutations of $\{1, \cdots, n\}$, scale appropriately, jitter if desired. Check to make sure that you didn't get the diagonal or something!



Dimension Reduction

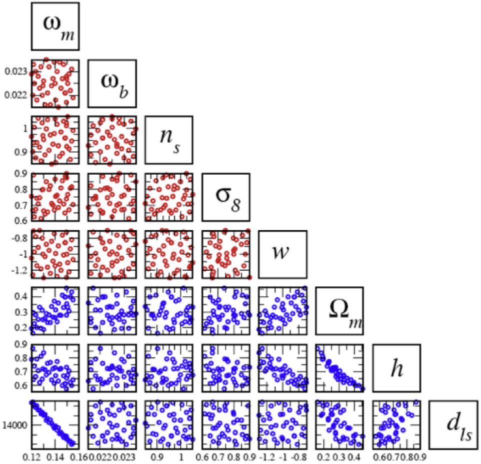
Often the output is multivariate (e.g. power spectrum).
Center, scale, and project onto a useful basis:

$$\begin{aligned}[y_1, \cdots, y_n] &\rightarrow [z_1, \cdots, z_n] \\ Z &= USV' \\ K &= US \\ w &= (K'K)^{-1}K'Z\end{aligned}$$

Often only need the weights for a handful of basis functions. Build GPs for the weights.

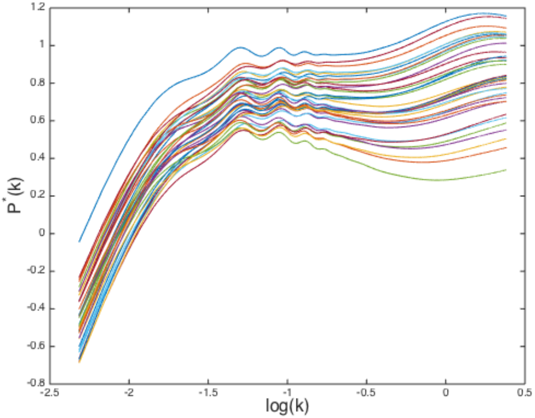
Los Alamos NATIONAL LABORATORY EST. 1943

Cosmic Emu

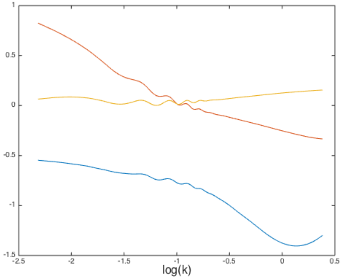


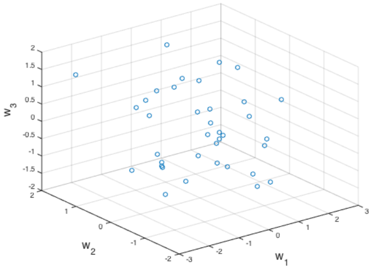
Los Alamos
NATIONAL LABORATORY
EST. 1943

Cosmic Emu

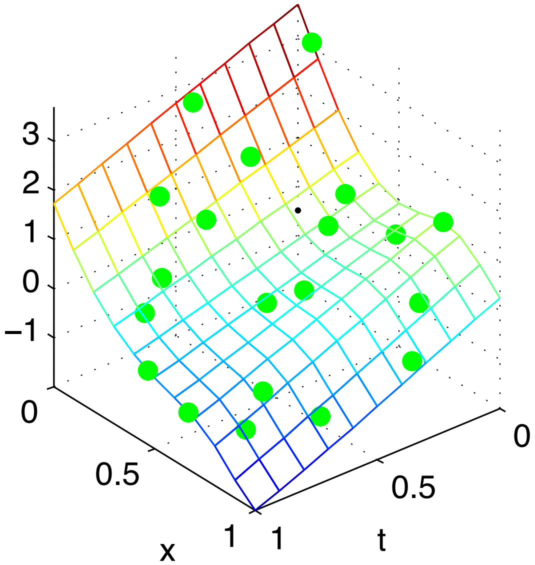


Cosmic Emu





Cosmic Emu



Cosmic Emu

Save the design, training weights, and the mean of the GP parameters. Use these to compute the conditional mean for the weights at new desired point. Multiply the saved basis vectors by the predicted weights.

Los Alamos NATIONAL LABORATORY EST. 1943

Discussion

In practice, the emulation and the parameter inference are done simultaneously. The posterior is messy, but it's just the product of a bunch of parts and MCMC stills gets the job done.