Ancient and modern ways to evaluate the posterior.

Douglas Nychka,
`www.image.ucar.edu/~nychka`

- Exact expressions
- Monte Carlo draws and creating an ensemble
- Markov Chain Monte Carlo MCMC
- Variational methods as a shortcut

Supported by the National Science Foundation *NCAR/IMAGe summer school Jul 2007*

Where we are

The posterior distribution is a distribution for the unknown parameters given the data and is proportional to the product of the likelihood and the prior:

$$[\theta|Y] \propto [Y|\theta][\theta]$$

The goal is to evaluate this density in θ.

Exact expressions

Simple example:

If $[Y|\theta] = N(\theta, 1)$ **and** $[\theta] = N(\mu, \sigma)$

then one can work out the normalization so that the posterior integrates to one.

Usually posteriors can be worked out exactly by using priors that are algebraically friendly with the likelihood — known as conjugate priors:

Gaussian-Gaussian, Binomial- Beta, Gamma- Inverse Gamma, Poisson-Gamma.

Products of Gaussians are Gaussian ....

Kalman filter/temperature example:

This is a Gaussian- Gaussian conjugate prior case.

$[Y|T] = N(HT, R)$ **and** $[T] = N(\mu, \Sigma)$

$[T|Y]$ **is also multivariate Gaussian given by the usual Kalman filter/Conditional Gaussian result.**

e.g posterior mean

$$\hat{T} = \mu + \Sigma H^T (H \Sigma H^T + R)^{-1} (Y - H\mu)$$

Problems with the Gaussian/Gaussian case

Computing:

Both the mean vector and the covariance matrix involve Σ and inverses of $H \Sigma H^T$ that may be difficult to evaluate due to the size of the T vector and large number of observations in Y.

Interpretation:

Even if you found Σ what would do with it?

How does it tell you about uncertainty in the estimated field?

Monte Carlo samples and ensembles

Use a random sample or ensemble from the posterior as a summary of the distribution.

E.g. $T_1, T_2, ...T_M$ are M random draws from $[T|Y]$
Each T_i is a whole coherent field and is equally likely given the data.

Posterior mean:

$\approx \frac{1}{M} \sum_{i=1}^{M} T_i$

More complicated features:

E.g. Regions of fields that exceed a particular value. Just use the frequency distribution based on the ensemble members.

Generating a sample from a Gaussian

Given a Gaussian posterior: $N(\widehat{T}, P_a)$**.**

find a square root matrix, A**, for** P_a**.**

$$P_a = AA^T$$

Draw from posterior

$$T = \widehat{T} + Au$$

u **is a vector of** $N(0,1)'s$**.**

A trick for the Kalman filter type problem

The random draw is similar to using the bootstrap

- **Generate a pseudo temperature field, T^* based on the prior. (use square root method)**
- **Generate pseudo data based on T^* and the likelihood.**

$$Y^* = HT^* + e^*$$

- **Use Kalman filter update equations to estimate the field based on the pseudo data.**

$$\hat{T}^* = \text{stuff}...Y^*...\text{etc.}$$

- **The draw:**

$$T = \hat{T} + (T^* - \hat{T}^*)$$

draw = posterior mean + (simulated error)

General ways to sample posterior.

Acceptance-rejection methods.

- Generate a draw from a *proposal* distribution. The proposal distribution is easy to simulate and chosen to be "close" to the posterior.
- Use a second algorithm to decide whether to accept or reject the proposal value.

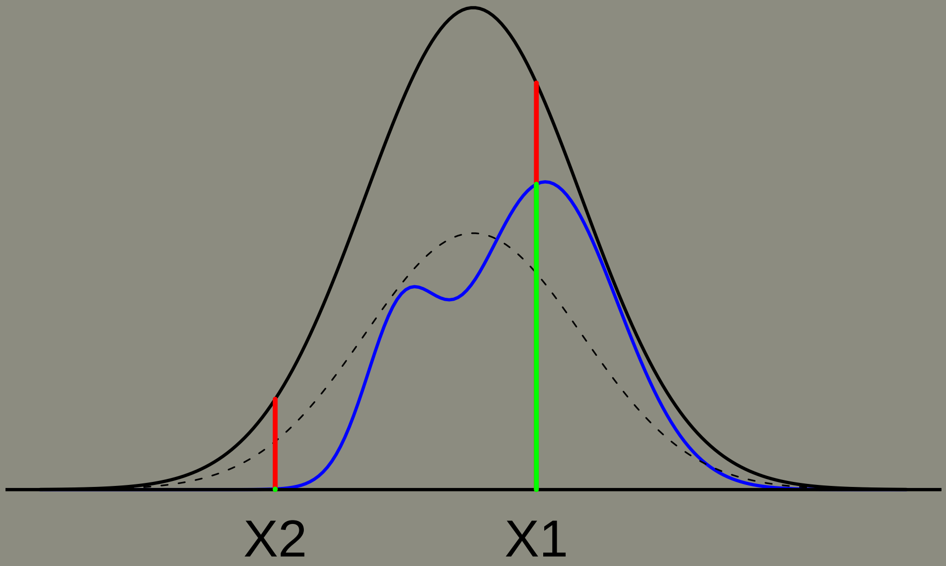
Repeat this procedure until you have enough acceptances

Acceptance sampling

Generate from **pdf** g to get a draw from **pdf** f

Assume that there is a bound K so that $0 < f(x) < Kg(x)$.

- Generate $X \sim g$
- **Accept** X with probability $f(X)/Kg(X)$



X1 is likely to be accepted and **X2** likely rejected.

Some comments

The better the proposal matches the target pdf the more often one accepts.

This method is hard to apply if we can't find the envelope for f!

Markov Chain Monte Carlo (MCMC)

Generate a random sequence $\{X_1, X_2, ...\}$*,*
a Markov chain, that will tend to be from f*.*

Proposal distribution can depend on the previous draw,

$g(x_2|x_1)$ **conditional pdf of proposal, given previous draw,**

Metroplis-Hastings Algorithm

- **Draw** X^* **from** $g(.|X_t)$ $R = \frac{f(X^*)g(X_t|X^*)}{f(X_t)g(X^*|X_t)}$
- **Accept** $X_{t+1} = X^*$ **with probability** R
  otherwise $X_{t+1} = X_t$**.**

The mysterious acceptance criterion, R

$$R = \frac{f(x^*)g(x|x^*)}{f(x)g(x^*|x)}$$

measures whether a proposal is as 'likely' as the current value.

Comments

We only need to know f up proportionally because it enters R as a ratio.

If all goes well the distribution of the $\{X_1, X_2, ...\}$ sequence will be from f.

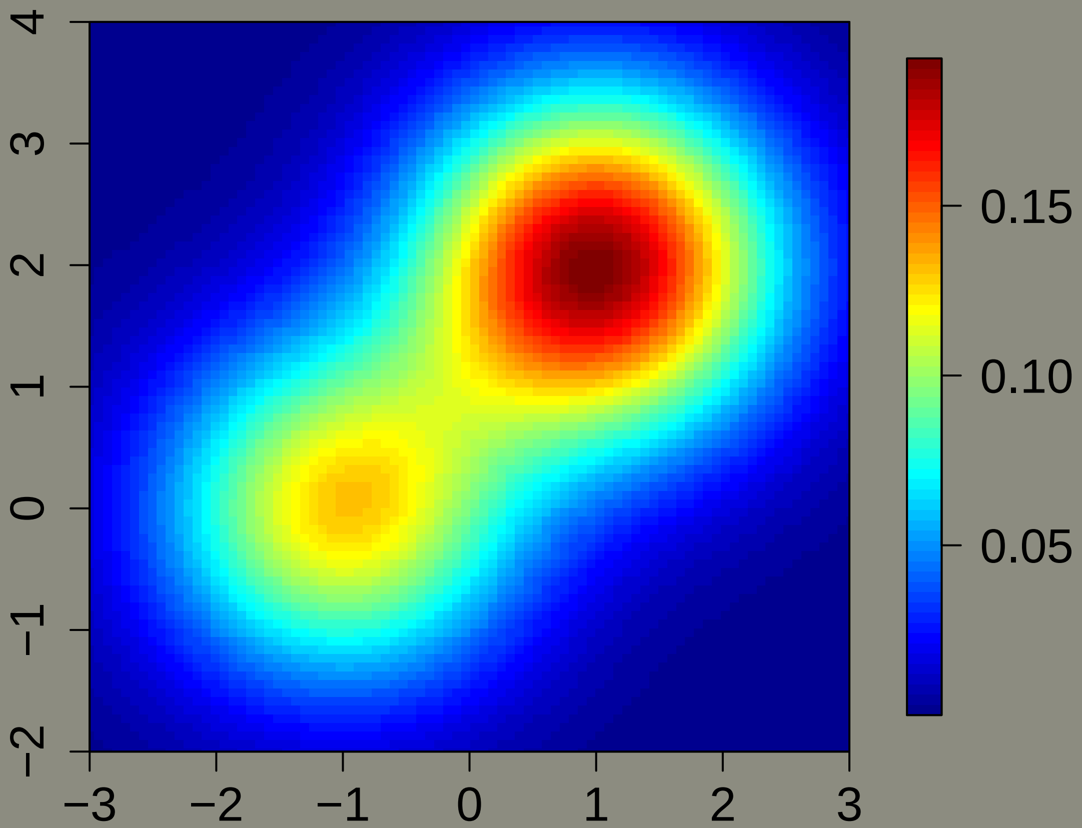
What can go wrong.

Proposals are not accepted very often.
(this is bad in many other contexts as well)

Sequence does not acheive a stationary state – gets stuck around particular values.

An example of MCMC

f is a bivariate distribution:

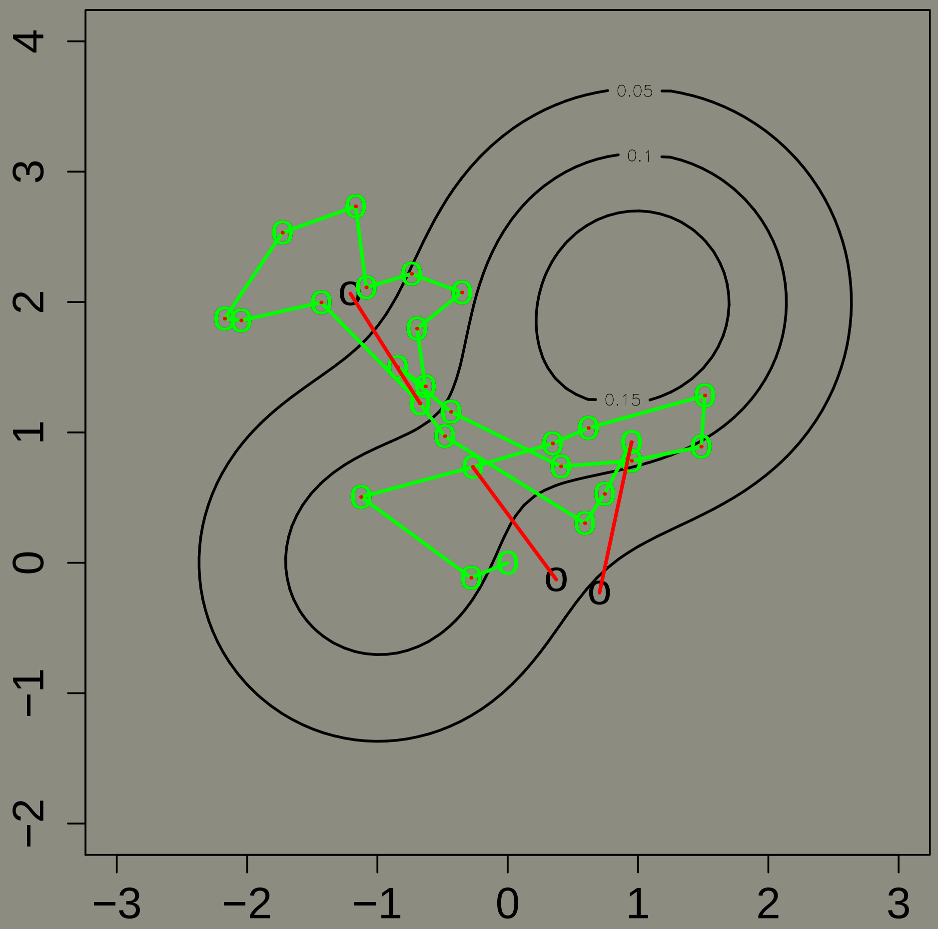


Generating a random walk proposal

Take the current vector and add a random Gaussian to each component.

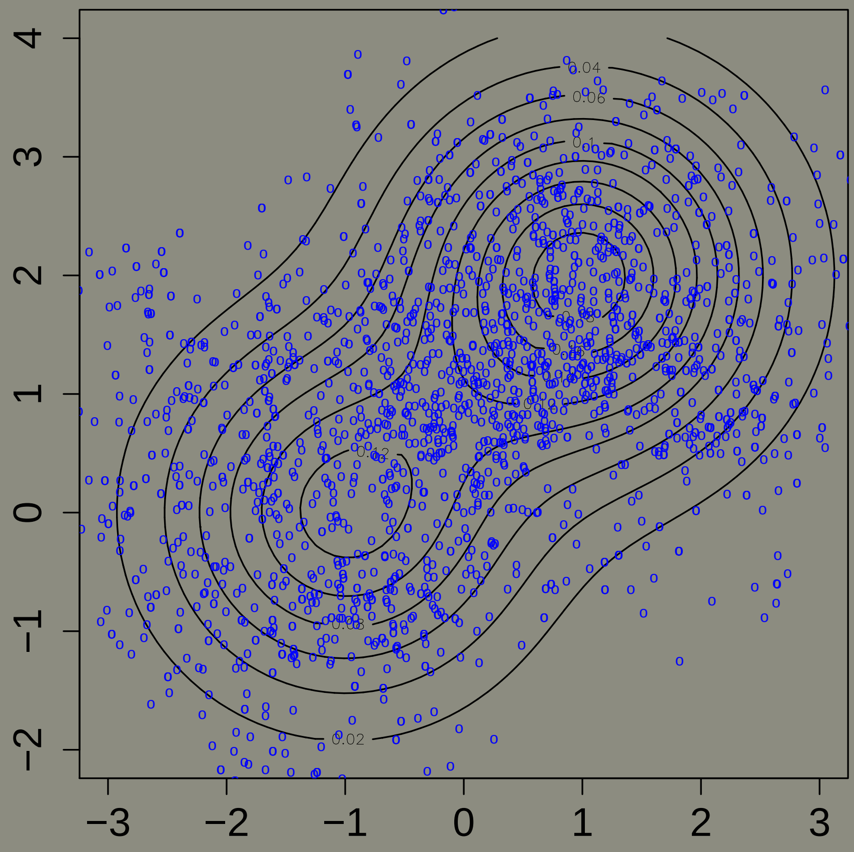
$$X^* = X_t + N(0, \sigma^2)$$

30 proposals $\sigma = .5$ $X_0 = (0, 0)$



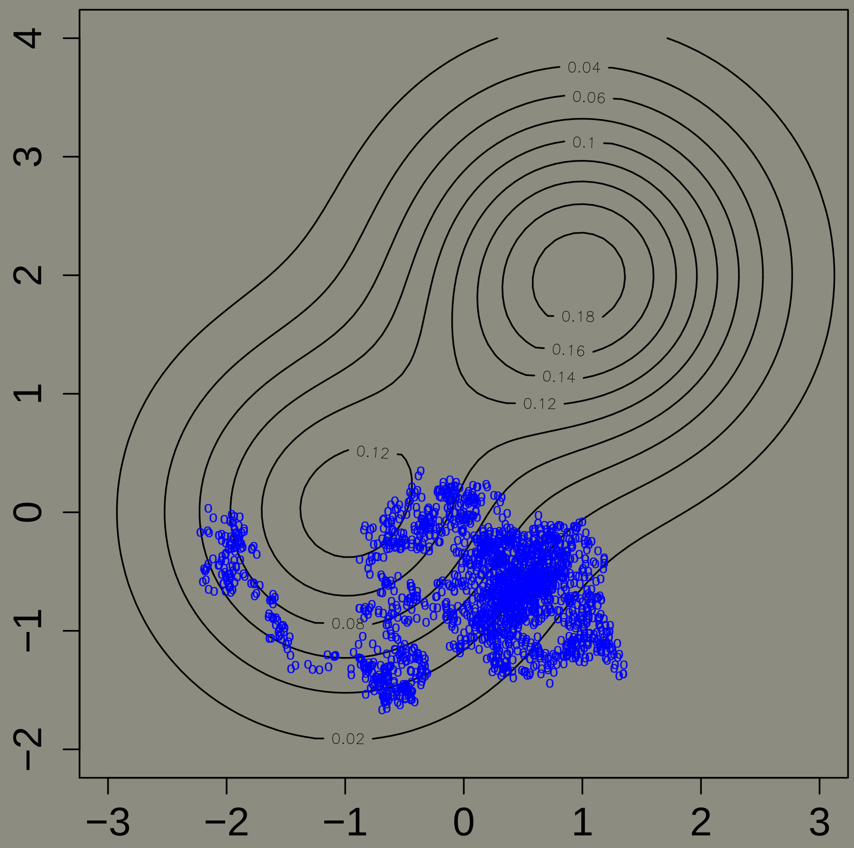
2000 proposals with $\sigma = .4$

About 85% of proposals are accepted.



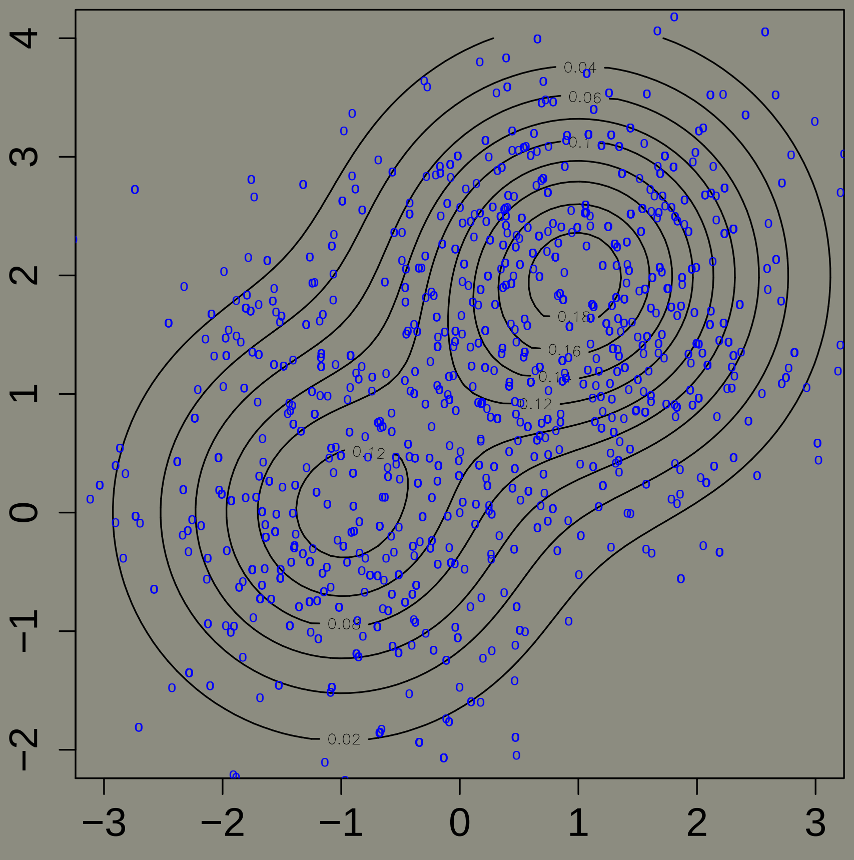
2000 proposals with $\sigma = .05$

About 97% of proposals are accepted.



2000 proposals with $\sigma = 2.0$

About 36% of proposals are accepted.



The big picture and a shortcut.

MCMC is a flexible procedure and usually easily to code.

It does need to be tuned to the problem and data at hand.

But when it converges one has an ensemble to represent the uncertainty of the parameters.

Given all the difficulties with sampling the posterior, finding the posterior mode may be a easy (and deterministic) way to get an estimate. This is the result of solving a variational problem.