PROC MCMC
Demonstration

NC State University, October 23, 2013

SAS | THE POWER TO KNOW.

Monte Carlo Methods

Monte Carlo methods

- involve the use of random sampling techniques based on computer simulation to obtain approximate solutions to integration problems

- have the aim of evaluating integrals or sums by simulation rather than exact or approximate analytic methods

- will be useful for Bayesian analysis to obtain posterior summaries from non-standard distributions.

Simulation Tools for High-Dimensional Sampling

- Metropolis and Metropolis-Hastings Algorithms
- Gibbs Sampler
- Adaptive Rejection Sampling Algorithm
- Independence Sampler
- Gamerman Algorithm

Markov Chain Convergence

- *Convergence* means that a Markov chain has reached its stationary (target) distribution.

- Assessing the Markov chain convergence is very important, as no valid inferences can be drawn if the chain is not converged.

- It is important to check the convergence for all the parameters and not just the ones of interest.

- Assessing convergence is a difficult task, as the chain converges to a distribution and not to a fixed point.

Diagnostic Plots – Good Mixing
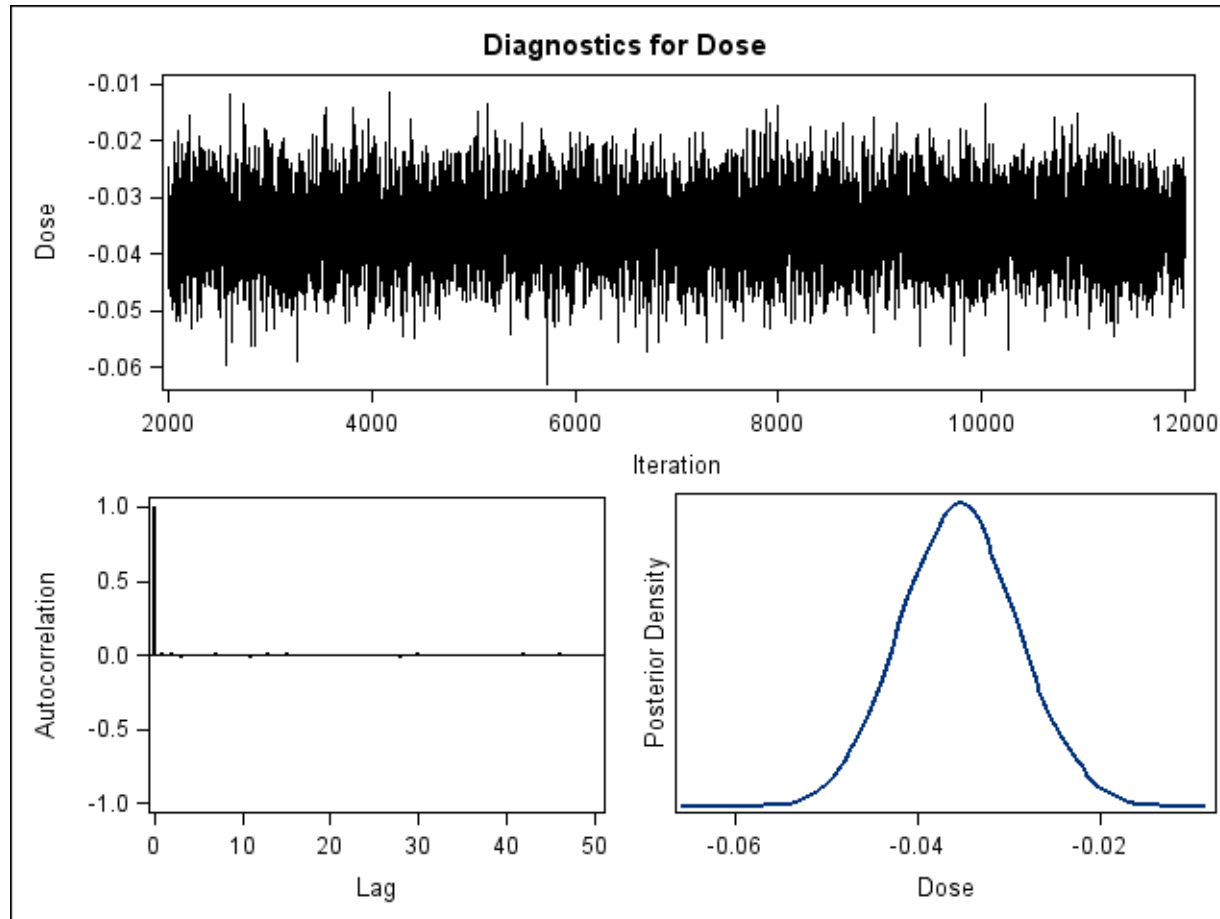
Diagnostic Plots – Poor Mixing
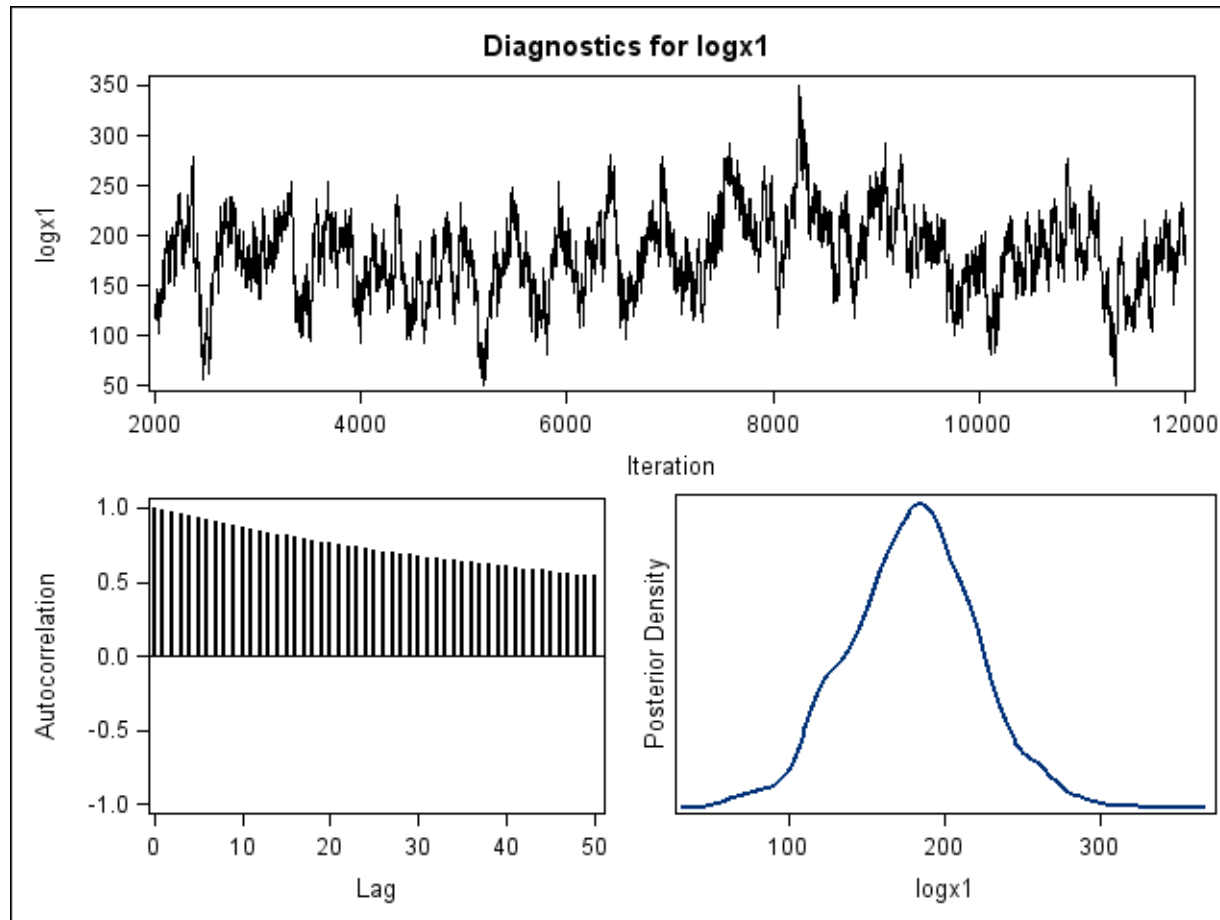
Gelman and Rubin Diagnostics

- This test uses multiple simulated MCMC chains with dispersed initial values and compares the variances within each chain and the variance between the chains.

- Large deviations between these two variances indicates non-convergence.

- A one-sided test based on a variance ratio test statistic is reported where large values indicate a failure to converge.

Geweke Diagnostics

- This tests whether the mean estimates have converged by comparing means from the early and latter part of the Markov chain.

- The test is a two-sided test based on a z-score statistic.

- Large absolute z values indicate a failure of convergence.

Heidelberger and Welch Diagnostics

These tests consist of two parts:

- a stationary portion test which assesses the stationarity of a Markov chain by testing the hypothesis that the chain comes from a covariance stationary process

- a half-width test which checks whether the Markov chain sample size is adequate to estimate the mean values accurately.

The stationary test

- is a one-sided test based on a Cramer-von Mises statistic.

The half-width test

- indicates non-convergence if the relative half-width statistic is greater than a predetermined accuracy measure.

Raftery and Lewis Diagnostics

- The test evaluates the accuracy of the estimated percentiles by reporting the number of samples needed to reach the desired accuracy of the percentiles.

- If the total number of samples needed are less than the Markov chain sample, the desired precision was not obtained.

- The test is specifically designed for the percentile of interest and does not provide information about convergence of the chain as a whole.

Effective Sample Size

Effective Sample Size

- is a measure of how well a Markov chain is mixing

- takes autocorrelation into account

- shows good mixing when it is close to the total sample size

Summary of Convergence Diagnostics

- There are no definitive tests of convergence.

- Visual inspection of the trace plots is often the most useful approach.

- Geweke and Heidelberger-Welch tests sometimes are statistically significant even when the trace plots look good.

- Oversensitivity to minor departures from stationarity does not impact inferences. Different convergence diagnostics are designed to protect you against different potential pitfalls.

Bayesian Analysis in SAS

Bayesian methods in SAS 9.3 are found in:

- the PHREG procedure, which performs regression analysis of survival data based on the Cox proportional hazards model

- the LIFEREG procedure, which fits parametric models to survival data

- the GENMOD procedure, which fits generalized linear models

- the MCMC procedure, which is a general purpose Markov chain Monte Carlo simulation procedure that is designed to fit Bayesian models.

The MCMC Procedure

PROC MCMC

- is a general purpose simulation procedure that uses Markov chain Monte Carlo (MCMC) techniques to fit a wide range of Bayesian models

- requires the specification of a likelihood function for the data and a prior distribution for the parameters

- enables you to analyze data that have any likelihood or prior distribution as long as they are programmable using SAS DATA step functions.

PROC MCMC Statements

- You declare the parameters in the model and assign the starting values for the Markov chain with the PARMS statements.

- You specify prior distributions for the parameters with the PRIOR statements.

- You specify the likelihood function for the data with the MODEL statements.

- The model specification is similar to PROC NLIN and shares much of the same syntax as PROC NLMIXED.

PROC MCMC Syntax

General form of the MCMC procedure:

PROC MCMC *options*;
    PARMS *parameters and starting values*;
    BEGINCNST;
      Programming Statements;
    ENDCNST;
    BEGINNODATA;
      Programming Statements;
    ENDNODATA;
    PRIOR *parameter ~ distribution*;
    MODEL *variable ~ distribution*;
    PREDDIST *<'label'>* OUTPRED=*SAS-data-set*
            <options>;
RUN;

Posterior Summaries

The posterior summaries include:

- Posterior mean, standard deviation, and percentiles
- Equal-tail and highest posterior density intervals
- Covariance and correlation matrices
- Deviance information criterion (DIC)