Bayesian Analysis in SAS/STAT® Software

Rich Bayesian Analysis

Overview

SAS/STAT software provides you three ways to conduct Bayesian analysis.

Specialized Models with Wide Application. The BGLIMM and BCHOICE procedures are high-performance procedures that are tailored to handle generalized linear mixed-effects models and discrete choice models, respectively.

Built in Bayesian Capabilities. The BAYES statement provides you Bayesian estimation to standard analyses available through existing procedures. This statement is suitable when you want quick and easy access to Bayesian analysis with minimal programming efforts.

General Bayesian Modeling Tool. The MCMC procedure provides you general purpose programming for Bayesian inference. You fit models in this procedure by specifying the priors and the likelihood function appropriate for your analysis.

The BGLIMM Procedure

The BGLIMM procedure is a sampling-based procedure that you use to obtain Bayesian inference for generalized linear mixed models (GLMMs) similar to the types of models you fit in the MIXED and GLIMMIX procedures. The procedure allows you to obtain inference from a class of robust models and minimizes the amount of syntax you need to provide to specify model attributes and prior distributions.

For a univariate response you can specify a normal distribution or choose from a variety of non-Gaussian distributions. For multivariate data you can use the multivariate normal distribution. The BGLIMM procedure allows you to build models for hierarchical nested or non-nested multilevel random effects and repeated measures data. It supports missing completely at random (MCAR) and missing at random (MAR) approaches in modeling missing data.

The Bayesian approach to GLMMs estimates the joint posterior disruption of all parameters in a model, including all fixed- and random-effect parameters. The Monte Carlo method numerically integrates out the random effects and propagates the uncertainties to the marginal posterior of the fixed-effects parameters. The BGLIMM procedure uses efficient Markov chain Monte Carlo (MCMC) sampling tools to estimate the posterior marginal distributions and stores them for further inference.

The BGLIMM procedure offers you many features including:

- syntax similar to other SAS/STAT regression procedures including the CLASS, MODEL, RANDOM, REPEATED, and ESTIMATE statements
- parallelized efficient sampling algorithms for fast performance.
- the ability to specify link functions and distributions for the response variable
- a suite of covariance structures for random effects and residuals
- built-in prior distributions for the regression coefficients, covariance parameters, and residuals
- the ability to output posterior samples to a data set for further inferences

The BGLIMM procedure is multithreaded and uses ODS Graphics to create graphs as part of its output. The table below shows the posterior summaries and intervals generated from PROC BGLIMM.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>50000</td>
<td>1.4840</td>
<td>0.3104</td>
<td>0.8872</td>
</tr>
<tr>
<td>Baseval</td>
<td>50000</td>
<td>0.5007</td>
<td>0.1026</td>
<td>0.3016</td>
</tr>
<tr>
<td>Drug a</td>
<td>50000</td>
<td>0.3237</td>
<td>0.0416</td>
<td>0.2424</td>
</tr>
<tr>
<td>Drug c</td>
<td>50000</td>
<td>0.5245</td>
<td>0.0420</td>
<td>0.4425</td>
</tr>
<tr>
<td>Drug p</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Built in Bayesian Capabilities

You have easy access to Bayesian analysis via the BAYES statement in the GENMOD, LIFEREG, PHREG, and FMM procedures of SAS/STAT software. Using these procedures you can perform Bayesian analysis for:

- generalized linear models, including logistic regression and Poisson regression
- Cox regression models
- frailty models
- accelerated failure time models
- piecewise exponential models
- finite mixture models
These procedures provide you:

- posterior point and interval statistics
- standard prior distributions
- a SAS data set that includes posterior samples for use in additional analyses

You assess convergence of the Markov chain using diagnostics such as the Gelman-Rubin, Geweke, Heidelberger-Welch, and Raftery-Lewis tests, along with diagnostic plots.

A trace plot (displayed below), an autocorrelation plot, and a posterior density plot are provided for each parameter that is estimated. Both posterior summaries and credible intervals are provided.

The MCMC procedure offers you a versatile RANDOM statement that can be used to specify:

- univariate or multivariate hierarchical models with arbitrary depth
- nested and non-nested models, nonnormal prior distributions
- categorical variables
- between-cluster covariance

You can write your own numerical functions and subroutines and carry out matrix-based computations. Additional embedded tools, such as a numerical integration function and ordinary differential solvers, add to the power of the procedure.

The MCMC procedure implements a large class of general MCMC algorithms (from Metropolis algorithm to Hamiltonian Monte Carlo) and assigns algorithms choices by default. You can change and select algorithms, specify blocking of parameters, and even write your own samplers to use in the procedure.

**The BCHOICE Procedure**

The BCHOICE procedure allows you to perform Bayesian analysis for discrete choice models. Discrete choice models are used in marketing research and microeconomics to model decision makers’ choices among alternative products and services. The collection of alternatives that are available to the decision makers is called a choice set.

The BCHOICE procedure allows you to specify the following types of models to get the expected choice probabilities of all the alternatives in a choice set:

- multinomial logit
- multinomial probit
- nested logit
- multinomial logit with random effects
- multinomial probit with random effects

You can specify prior distributions for the regression coefficients and the covariance of the random effects. Otherwise, the BCHOICE procedure will use non-informative priors.

PROC BCHOICE obtains samples from the corresponding posterior distributions and posterior predictive distribution of the choice probabilities, produces summary and diagnostic statistics, and saves the posterior samples in an output data set that can be used for further analysis.

**For More Information**

For complete information about all SAS/STAT procedures, see the documentation available at support.sas.com/statistics/.