PharmaSUG 2013 - Paper BB12

Bayesian Analysis of Survival Data with SAS PHREG Procedure

Ryan Brady, Texas A&M, College Station, Tx

ABSTRACT

Bayesian analysis has advantages in flexibility and ease of interpretation, but is mathematically complex and computationally intense. Fortunately, the SAS BAYES statement obscures much of the complexity, allowing statisticians and programmers to easily take advantage of that increased flexibility. The BAYES statement is available for four procedures, including PHREG, which is used for analysis of survival data using the Cox proportional hazards model. PROC PHREG uses the partial likelihood as the likelihood, and uses a MCMC Gibbs sampler to generate a posterior distribution. Diagnostic plots are available to determine sampler properties such as mixing, convergence and stationarity.

INTRODUCTION

BAYESIAN STATISTICS

Baysian statistics is a method of updating existing beliefs based on new data. It is an extension of Bayes theorem, which, can be written thus:

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

Where θ refers to the value of some parameter of interest, and Y refers to the observed evidence. $P(Y|\theta)$ is the likelihood function, $P(\theta)$ is referred to as the prior, $P(\theta|Y)$ is the posterior probability, and $P(Y)$ is typically treated as some constant of proportionality.

The use of a prior is the key advantage of Bayesian statistics. The prior may be noninformative or skeptical, indicating that there is no information to predict the actual value of the parameter, it may be based on prior information, or it may be constructed in order to limit the distribution to only plausible values (such as positive reals). The choice of prior can be somewhat of an art, and in SAS 9.3 priors for the log-hazards are limited to the uniform and normal priors.

Bayesian analysis can be requested in SAS software with the BAYES statement in select procedures.

BAYESIAN STATISTICS IN CLINICAL TRIALS

It may sometimes appear that the primary difference between classical frequentist statistics and Bayesian statistics is philosophical. For both the businesses and patients involved in clinical trials, however, the benefits of using Bayesian statistics are very real. For the businesses conducting clinical trials, Bayesian statistics can be used to construct adaptive clinical trials with a lower average sample number and higher power than classical methods, incorporating efficacy, futility and cost. For patients, this can translate to quality of life improvements in the form a more timely termination for ineffective treatments.

On February 5, 2010, FDA issued a guidance document entitled “Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials.” This document recognizes the benefits of Bayesian statistics in clinical trials both through the use of incorporating prior information (preferably based on previous empirical studies), and in the creation

of adaptive trials. Allowing Bayesian clinical trials is specifically justified by Section 513(a)(3) of the Federal Food, Drug, and Cosmetic Act which mandates that FDA shall consider “the least burdensome appropriate means of evaluating effectiveness of a device that would have a reasonable likelihood of resulting in approval” (21 U.S.C. 360c(a)(3)).

COX PROPORTIONAL HAZARDS MODEL

The Cox proportional hazards model is an approach to the analysis of survival data which examines the relative hazard rates for a condition between covariates. A common application is the relative hazard between patients given a treatment and patients given a placebo. The Cox proportional hazards model is called for in SAS software by the PHREG procedure.

MARKOV CHAIN MONTE CARLO (MCMC) GIBBS SAMPLING

SAS software uses a Markov chain Monte Carlo method known as Gibbs sampling to simulate posterior distributions. The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, and samples each parameter from the full conditional distribution (conditional on the data and all other parameters), using the most recent values for each other parameter. Deriving the full conditional distributions can be somewhat challenging; fortunately, the SAS software obscures these mathematical steps, allowing the statistician to more easily approach Bayesian analysis.

DIGITALIS INVESTIGATION GROUP DIGOXIN TRIAL

The following analysis was performed using permuted data from the Digitalis Investigation Group (DIG) trial, and is presented for the purpose of demonstrating the BAYES statement in PROC PHREG. Multidimensional relationships may not have been preserved in the permutation process, so the analyses here should not be used to draw any conclusions about the DIG study.

BAYESIAN ANALYSIS OF SURVIVAL DATA WITH PROC PHREG

ANATOMY OF THE BAYES STATEMENT

The BAYES statement calls for Bayesian analysis in the PHREG procedure. The options for the BAYES statement belong to four general categories.

- Monte Carlo simulation parameters
- Model and prior options
- Diagnostics and summaries
- Saving the posterior data

Monte Carlo Simulation Parameters

Monte Carlo simulation is a random process, so the SEED option is necessary for reproducibility. In general, there isn’t any reason to use a value other than SEED=1.

Likewise, there is rarely any reason to alter the INITIAL or THINNING options. The initial values for Monte Carlo simulation are, in most cases, somewhat inconsequential. Assuming that the simulation has good mixing (the ability for the current state to quickly move to another state), the initial values will quickly give way to more realistic parameter values. In the case of a multimodal posterior distribution, however, it may be necessary to run the simulation with a variety of different initial values.

The THINNING option is used to only retain every *n*th simulated sample, purportedly to reduce autocorrelation. However, this is rarely, if ever advisable. To obtain a sample with 10000 observations, with THINNING=8, the

software must generate 80000 observations. Given that autocorrelated samples are only unrepresentative in the short run, it is probable that a simulation with 80000 observations will be more precise than one thinned from 80000 observations to only 10000. In other words, it is almost always preferable to simply increase the number of MCMC iterations if autocorrelation is a concern. (Link and Eaton, 2012)

The NBI and NMC options are, in contrast, quite important. The NBI option controls the number of iterations that are discarded as “burn in” before the simulation is assumed to have reached stationarity (a noisy state that appears to have no patterns). Setting NBI too high has no effect other than an increase in execution time, but setting NBI too low may include a portion of the simulation before it has reached stationarity. If this is the case, the posterior distribution will not be truly representative. The default setting of 2000 will be sufficient in most cases.

The NMC option controls the number of iterations after burn-in. The default number of 10000 iterations is likely sufficient for more applications. If there is poor mixing or high autocorrelation it may be necessary to increase the number of MCMC iterations to compensate. Conversely, if the simulation is well behaved it may be acceptable to reduce the number of iterations in order to improve execution time (particularly when testing code not intended to produce a final report).

Model and Prior Options

The COEFF option is used to specify the prior distribution, from either the default flat uniform distribution or a normal distribution. The normal prior can either be set to user-specified means and variances for each variable using the INPUT option with a SAS data set, or left at the default with mean 0 and variance 10^6 for each variable.

The PIECEWISE option may be used to instead specify the piecewise constant baseline function. The use of this option is outside the scope of this paper, but it should be noted that the intervals add to the number of variables that must be simulated in the Gibbs sampler, leading to a significant increase in execution time. For example, with the following SAS code, the default proportional hazards model had an execution time of 45 seconds, and the piecewise model with the default 8 intervals had an execution time of 4 minutes.

```
PROC PHREG data=Dig;
class TRTMT;
model  DWHFDAYS*DWHF(0)= TRTMT;
       bayes seed=1 piecewise;
       hazardratio TRTMT;
run;

PROC PHREG data=Dig;
class TRTMT;
model  DWHFDAYS*DWHF(0)= TRTMT;
       bayes seed=1;
       hazardratio TRTMT;
run;
```

Diagnostics

Diagnostics for the MCMC simulation come in two forms: diagnostic tests specified by the DIAG option, or diagnostic plots specified by the PLOTS option.

Convergence of the MCMC simulation can be determined through the Gelman-Rubin test (GELMAN keyword) or the Geweke test (GEWEKE keyword). The Gelman-Rubin test runs multiple simulations (three, by default) and compares the results. For this reason, using the Gelman-Rubin test greatly increases execution time. The Geweke test instead compares a fraction of the first part of the post burn-in Markov chain, with a fraction of the last part of the chain. Because multiple simulations are not required, the Geweke test does not significantly increase execution time. Convergence can also be visually assessed through the trace plot (TRACE keyword). A trace which maintains constant mean and variance has most likely converged to stationarity.

Autocorrelation is assessed by the autocorrelation diagnostic (AUTOCORR keyword) or the autocorrelation plot (AUTOCORR keyword). As previously stated, autocorrelation does not necessarily indicate a problem with the analysis, but may indicate that a longer simulation is required. Autocorrelation is closely related to mixing. A trace that indicates poor mixing will also exhibit stronger autocorrelation. The effect of autocorrelation on effective sample size can be seen with the ESS keyword for the DIAG option.

Two additional tests, the Rafferty-Lewis diagnostic and the Heidelberger-Welch Diagnostics can be used to determine whether or not the number of Markov chain iterations is sufficient to meet the desired accuracy. These are called by the RAFFERTY keyword and the HEIDELBERGER keyword for the DIAG option. Their use is outside the scope of this document.

Posterior Sample

The posterior sample can be either handled automatically, or can be saved to a new SAS dataset using the OUT option. The OUT option produces a new data set which contains the post burn-in MCMC values for each parameter, and two additional columns for iteration number and log likelihood.

Automatic analysis of the posterior samples is achieved through the use of the DENSITY keyword to the PLOTS option and the STATISTICS option. The default settings for the STATISTICS option return posterior summaries (mean, standard deviation, and 25%/50%/75% quantiles) as well as the equal tail and highest probability density intervals at the 0.05 level. Correlation and covariance matrices are also available through the CORR and COV keywords.

ANALYSIS OF PERMUTED DIXOGIN DATA

DATA PREPARATION AND MODEL SELECTION

The ‘import file’ functionality in SAS 9.3 was used to import the data as a comma separated value (.csv) file with default variable names from column headers. Parameters for Bayesian analysis were selected using the SELECTION option in the MODEL statement of the PHREG procedure with all potential predictors and no interactions. The BAYES statement is not available in conjunction with the SELECTION option. (If it were, the execution time would likely be unbearably long). The use of this procedure is outside of the scope of this paper, but the resulting model is presented below. There were multiple outcomes recorded in the data, but only DWHF (death from worsening heart failure) is presented here.

SAS CODE

Classical Analysis:

```
PROC PHREG data=dig;
class TRTMT RACE SEX FUNCTCLS CHFETIOL DIURETK KSUPP
      HYDRAL VASOD;
model  DWHFDAYS*DWHF(0)=
            CHESTX FUNCTCLS DIURET EJF_PER DIGUSE TRTMT
            KSUPP HYDRAL  SEX RACE CHFETIOL DIURETK ANGINA
            CHFDUR HEARTRTE;
      hazardratio TRTMT;
run;
```

Bayesian Analysis:

```
PROC PHREG data=dig;
class TRTMT RACE SEX FUNCTCLS CHFETIOL PREVMI ANGINA
      DIABETES HYPERTEN DIGUSE DIURETK DIURET KSUPP
      ACEINHIB NITRATES HYDRAL VASOD;
model  DWHFDAYS*DWHF(0)=
            CHESTX FUNCTCLS DIURET EJF_PER DIGUSE TRTMT
            KSUPP HYDRAL  SEX RACE CHFETIOL DIURETK ANGINA
            CHFDUR HEARTRTE;
      bayes seed=1 diagnostic=all
       plots=all;
      hazardratio TRTMT;
run;
```

Note: The output from this SAS code is extensive, and is not presented here in its entirety.

RUN TIME

Run time will vary given differences in hardware. On a particular machine however, the following CPU times were observed.

- Classical analysis CPU Time: 0.28 sec.
- Bayesian analysis CPU Time with `diagnostic=(autocorr ess)`, and `plots=trace`: 1 hr 19 min 6.79 sec.
- Bayesian analysis CPU Time with all plots and diagnostics: 3 hr 55 min 18.96 sec.

Statisticians and programmers would be well advised to be aware of the potential length of calculation for Bayesian approaches, and plan accordingly.

OUTPUT

For the sake of brevity, the output is abridged, and only select variables are chosen for each section of the output.

MCMC Settings

Because a prior distribution was not specified, the software used classical methods to select initial values for the Markov chain. This is easy to see by comparing the classical estimates with the initial Markov chain values:

Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**95% Confidence Limits**	
CHESTX	1	2.8117	0.3202	2.1840	3.4394
FUNCTCLS1	1	-1.0017	0.1428	-1.2816	-0.7218
FUNCTCLS2	1	-0.7244	0.1257	-0.9707	-0.4781

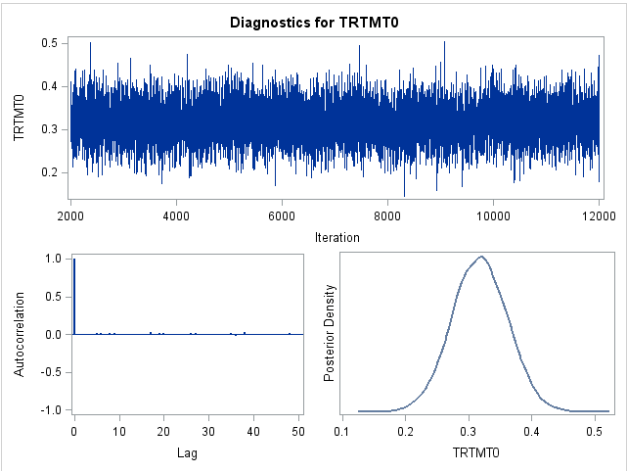
Figure 1. Excerpt of MLE Output

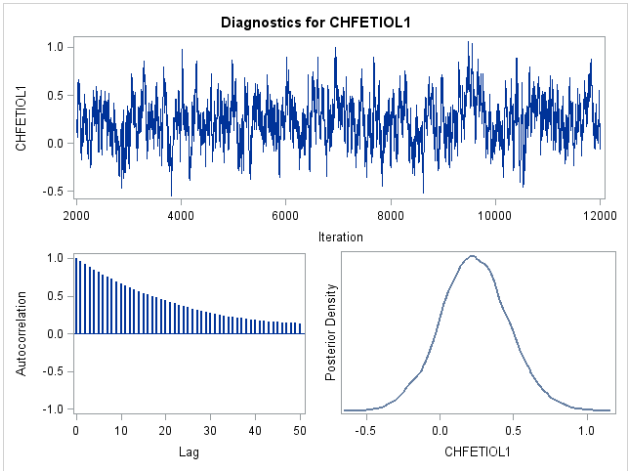
Chain	Seed	CHESTX	FUNCTCLS1	FUNCTCLS2
1	1	2.8117	-1.0017	-0.7244
2		1.851	-1.4301	-1.1014
3		3.7724	-0.5732	-0.3474

Figure 2. Excerpt of MCMC Initial Values

Markov Chain Diagnostics

Because DIAGNOSTIC=ALL and PLOTS=all were called, there are multiple diagnostics to choose from. The two trace plots below illustrate good mixing (figure 1) and poor mixing (figure 2). Likewise, there is little autocorrelation for the simulation of TRTMT0 and high autocorelation for the simulation of CHFETIOL1. In both cases, the Markov Chain has converged. When initial chain values are based on the maximum likelihood estimates, rapid convergence is generally expected.





A similar assessment can be made from the diagnostic tests.

Gelman-Rubin Diagnostics		
Parameter	**Estimate**	**97.5% Bound**
TRTMT0	1.0001	1.0004
CHFETIOL1	1.0155	1.0548

Figure 5. Excerpt of Gelman-Rubin Diagnostic Output

The Gelman-Rubin diagnostic shows both parameters with an estimate close to 1, indicating convergence.

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
TRTMT0	2.6524	0.008

Figure 3. Plots for Parameter TRMT0

Figure 4. Plots for Parameter CHFETIOL1

CHFETIOL1	-2.3342	0.0196

Figure 6. Excerpt of Geweke Diagnostic Output

The Geweke test is a two sided test based on a z-score statistic. The interpretation of the score is familiar – the small p-value indicates that the two fractions have similar means.

Effective Sample Sizes			
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
TRTMT0	10000	1	1
CHFETIOL1	195.6	51.1361	0.0196

Figure 7. Excerpt of Effective Sample Size (ESS) Output.

The poor mixing and consequent high autocorrelation in the CHFETIOL1 variable suggests that it may be necessary to run a longer simulation if a higher effective sample size is desired.

Hazard Ratios

The BAYES option automatically applies to the request for hazard ratios, generating the mean, standard deviation, quantiles, equal tail interval, and highest probability density interval for the requested hazard ratios. These can be obtained manually from the estimates of posterior means, but using the HAZARD statement is much more convenient.

Hazard Ratios for TRTMT			
Description	**Point Estimate**	**95% Wald Confidence Limits**	
TRTMT 0 vs 1	1.374	1.260	1.499

Figure 8. Hazard Ratio Output from Classical Analysis

Hazard Ratios for TRTMT										
				Quantiles						
Description	**N**	**Mean**	**Standard Deviation**	**25%**	**50%**	**75%**	**95% Equal-Tail Interval**		**95% HPD Interval**	
TRTMT 0 vs 1	10000	1.3753	0.0604	1.3338	1.3740	1.4154	1.2609	1.4958	1.2592	1.4938

Figure 9. Hazard Ratio Output from Bayesian Analysis.

DISCUSSION

Bayesian analysis can be a powerful tool in clinical research, and with the FDA's guidance document on the use of these methodologies, it is becoming increasingly important for statistical programmers to understand and implement them. Despite the continued increases in computing power, programmers should also keep mind of execution time concerns when performing Bayesian analysis, avoiding extraneous parameters and excessively long simulations.

CONTACT

Comments and questions are welcome. Please address all correspondence to:

Ryan Brady
rdbrady@gmail.com
1117 Guerrero St.
San Francisco, CA 94110
(415) 883-8400

REFERENCES

- Berry, D. (2004). "Bayesian Statistics and the Efficiency and Ethics of Clinical Trials" Statistical Science, Vol. 19, No. 1, 175–187 doi: 10.1214/088342304000000044
- Cheng, Y. and Shen, Y. (2005) "Bayesian Adaptive Designs for Clinical Trials" Biometrika Vol. 92, No. 3, 633-646
- Link, W. A. & Eaton, M. J. (2012) "On thinning of chains in MCMC" Methods in Ecology and Evolution Vol. 3 No. 1, 112-115 doi: 10.1111/j.2041-210X.2011.00131.x
- "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," Food and Drug Administration
- SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.
- Digitalis Investigation Group data set (https://biolincc.nhlbi.nih.gov/studies/dig/)