Paper SAS1775-2015

An Introduction to Bayesian Analysis with SAS/STAT® Software

Maura Stokes, Fang Chen, and Funda Gunes
SAS Institute Inc.

Abstract

The use of Bayesian methods has become increasingly popular in modern statistical analysis, with applications in numerous scientific fields. In recent releases, SAS® has provided a wealth of tools for Bayesian analysis, with convenient access through several popular procedures as well as the MCMC procedure, which is designed for general Bayesian modeling. This paper introduces the principles of Bayesian inference and reviews the steps in a Bayesian analysis. It then describes the built-in Bayesian capabilities provided in SAS/STAT®, which became available for all platforms with SAS/STAT 9.3, with examples from the GENMOD and PHREG procedures. How to specify prior distributions, evaluate convergence diagnostics, and interpret the posterior summary statistics is discussed.

Foundations

Bayesian methods have become a staple for the practicing statistician. SAS provides convenient tools for applying these methods, including built-in capabilities in the GENMOD, FMM, LIFEREG, and PHREG procedures (called the built-in Bayesian procedures), and a general Bayesian modeling tool in the MCMC procedure. In addition, SAS/STAT 13.1 introduced the BCHOICE procedure, which performs Bayesian choice modeling. With such convenient access, more statisticians are digging in to learn more about these methods.

The essence of Bayesian analysis is using probabilities that are conditional on data to express beliefs about unknown quantities. The Bayesian approach also incorporates past knowledge into the analysis, and so it can be viewed as the updating of prior beliefs with current data. Bayesian methods are derived from the application of Bayes' theorem, which was developed by Thomas Bayes in the 1700s as an outgrowth of his interest in inverse probabilities.

For events A and B, Bayes' theorem is expressed as

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

It can also be written as

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|\bar{A})\Pr(\bar{A})}$$

where $\bar{A}$ means *not* A. If you think of A as a parameter θ and B as data y, then you have

$$\Pr(\theta|y) = \frac{\Pr(y|\theta)\Pr(\theta)}{\Pr(y)} = \frac{\Pr(y|\theta)\Pr(\theta)}{\Pr(y|\theta)\Pr(\theta) + \Pr(y|\bar{\theta})\Pr(\bar{\theta})}$$

The quantity $\Pr(y)$ is the marginal probability, and it serves as a normalizing constant to ensure that the probabilities add up to unity. Because $\Pr(y)$ is a constant, you can ignore it and write

$$\Pr(\theta|y) \propto \Pr(y|\theta)\Pr(\theta)$$

Thus, the likelihood $\Pr(y|\theta)$ is being updated with the prior $\Pr(\theta)$ to form the posterior distribution $\Pr(\theta|y)$.

For a basic example of how you might update a set of beliefs with new data, consider a situation where researchers screen for vision problems in children in an after-school program. A study of 14 students chosen at random produces two students with vision issues. The likelihood is obtained from the binomial distribution:

$$L = \binom{14}{2} p^2(1-p)^{12}$$

Suppose that the parameter p only takes values { 0.1, 0.12, 0.14, 0.16, 0.18, 0.20}. Researchers have prior beliefs about the probabilities of these values, and they assign them prior weights. Columns 1 and 2 in Table 1 contain the possible values for p and the prior probability weights, respectively. You can then compute the likelihoods for each of the values for p based on the study results, and then you can weight them with the corresponding prior weight. Column 5 contains the posterior values, which are the computed values displayed in column 4 divided by the normalizing constant 0.2501. Thus, the prior beliefs have been updated to a posterior distribution by accounting for the data obtained by the study. The posterior values are similar to, but different from, the likelihood.

Table 1 Empirical Posterior Distribution

p	Prior Weight	Likelihood	Prior x Likelihood	Posterior
0.10	0.10	0.257	0.0257	0.103
0.12	0.15	0.2827	0.0424	0.170
0.14	0.20	0.2920	0.0584	0.233
0.16	0.20	0.2875	0.0576	0.230
0.18	0.15	0.2725	0.0410	0.164
0.20	0.10	.2501	0.0250	0.100
Total	1		.2501	1

In a nutshell, this is what any Bayesian analysis does: it updates your beliefs about the parameters by accounting for additional data. You weight the likelihood for the data with the prior distribution to produce the posterior distribution. If you want to estimate a parameter θ from data $\mathbf{y} = \{y_1, \ldots, y_n\}$ by using a statistical model described by density $p(\mathbf{y}|\theta)$, Bayesian philosophy says that you can't determine θ exactly but you can describe the uncertainty by using probability statements and distributions. You formulate a prior distribution $\pi(\theta)$ to express your beliefs about θ. You then update those beliefs by combining the information from the prior distribution and the data, described with the statistical model $p(\theta|\mathbf{y})$, to generate the posterior distribution $p(\theta|\mathbf{y})$.

$$p(\theta|\mathbf{y}) = \frac{p(\theta, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{\int p(\mathbf{y}|\theta)\pi(\theta)d\theta}$$

The quantity $p(\mathbf{y})$ is the normalizing constant of the posterior distribution. It is also called the marginal distribution, and it is often ignored, as long as it is finite. Hence $p(\theta|\mathbf{y})$ is often written as

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)\pi(\theta) = L(\theta)\pi(\theta)$$

where L is the likelihood and is defined as any function that is proportional to $p(\mathbf{y}|\theta)$. This expression makes it clear that you are effectively weighting your likelihood with the prior distribution. Depending on the influence of the prior, your previous beliefs can impact the generated posterior distribution either strongly (subjective prior) or minimally (objective or noninformative prior).

Consider the vision example again. Say you want to perform a Bayesian analysis where you assume a flat prior for p, or one that effectively will have no influence. A typical flat prior is the uniform,

$$\pi(p) = 1$$

and because the likelihood for the binomial distribution is written as

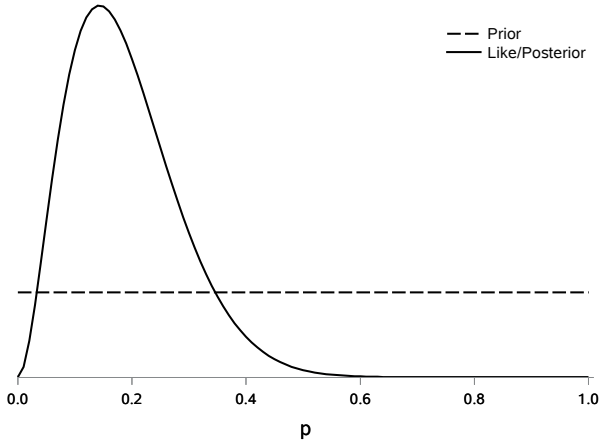
$$L(p) = \binom{n}{y} p^y (1-p)^{n-y}$$

you can write the posterior distribution as

$$\pi(p|y) \propto p^2(1-p)^{12}$$

which is also a beta (3,13) distribution. The flat prior weights equally on the likelihood, making the posterior distribution have the same functional form as the likelihood function. The difference is that, in the likelihood function, the random variable is y; in the posterior, the random variable is p. Figure 1 displays how the posterior distribution and the likelihood have the same form for a flat prior.

Figure 1 Beta (3,13) Posterior with Flat Uniform Prior



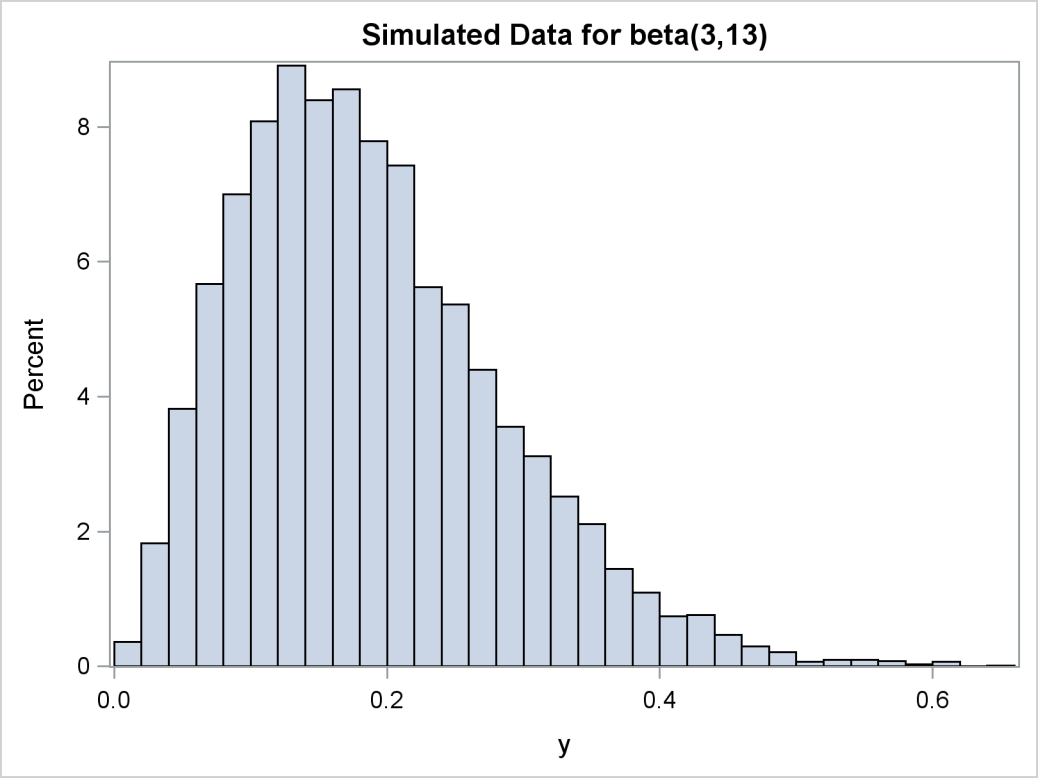
You can compute some summary measures of the posterior distribution directly such as an estimate of the mean of p and its variance, but you might want to compute other measures that aren't so straightforward, such as the probability that p is greater than a certain value such as 0.4. You can always simulate data from the beta distribution and address such questions by working directly with the simulated samples.

The following SAS statements create such a simulated data set for the beta (3,13) distribution:

```
data seebeta;
   %let N =10000;
   call streaminit (1234);
   a= 3; b= 13;
   do i=1 to &N;
     y = rand("beta", a, b );
   output;
   end;
run;
```

The results can be seen by the histogram in Figure 2 generated by using with the SGPLOT procedure.

Figure 2 Simulation for beta(3,13)



The mass of the distribution lies between 0.0 and 0.4, with the heaviest concentration between 0.1 and 0.2. Very little of the distribution lies beyond $p = 0.5$. If you want to determine the probability of $p > 0.4$, you would total the area under the curve for $p > 0.4$.

More often, closed forms for the posterior distribution like the beta distribution discussed above are not available, and you have to use simulation-based methods to estimate the posterior distribution itself, not just draw samples from it for convenience. Thus, the widespread use of Bayesian methods had to wait for the computing advances of the late 20th century. These methods involve repeatedly drawing samples from a target distribution and using the resulting samples to empirically approximate the posterior distribution. Markov chain Monte Carlo (MCMC) methods are used extensively. A Markov chain is a stochastic process that generates conditional independent samples according to a target distribution; Monte Carlo is a numerical integration technique that finds an expectation of an integral. Put together, MCMC methods generate a sequence of dependent samples from the target posterior distribution and compute posterior quantities of interest by using Monte Carlo. Popular and flexible MCMC simulation tools are the Metropolis, Metropolis-Hastings, and Gibbs sampling algorithms as well as numerous variations.

This paper does not discuss the details of these computational methods, but you can find a summary in the "Introduction to Bayesian Analysis" chapter in the *SAS/STAT User's Guide* as well as many references. However, understanding the need to check for the convergence of the Markov chains is essential in performing Bayesian analysis, and this is discussed later.

The Bayesian Method

Bayesian analysis is all about the posterior distribution. Parameters are random quantities that have distributions, as opposed to the fixed model parameters of classical statistics. All of the statistical inferences of a Bayesian analysis come from summary measures of the posterior distribution, such as point and interval estimates. For example, the mean or median of a posterior distribution provides point estimates for θ, whereas its quantiles provide *credible intervals.*

These credible intervals, also known as credible sets, are analogous to the confidence intervals in frequentist analysis. There are two types of credible intervals: the equal-tail credible interval describes the region

between the cut-points for the equal tails that has $100(1-\alpha)\%$ mass, while the highest posterior density (HPD), is the region where the posterior probability of the region is $100(1-\alpha)\%$ and the minimum density of any point in that region is equal to or larger than the density of any point outside that region. Some statisticians prefer the equal-tail interval because it is invariant under transformations. Other statisticians prefer the HPD interval because it is the smallest interval, and it is more frequently used.

The prior distribution is a mechanism that enables the statistician to incorporate known information into the analysis and to combine that information with that provided by the observed data. For example, you might have expert opinion or historical information from previous studies. You might know the range of values for a particular parameter for biological reasons. Clearly, the chosen prior distribution can have a tremendous impact on the results of an analysis, and it must be chosen wisely. The necessity of choosing priors, and its inherent subjectivity, is the basis for some criticism of Bayesian methods.

The Bayesian approach, with its emphasis on probabilities, does provide a more intuitive framework for explaining the results of an analysis. For example, you can make direct probability statements about parameters, such as that a particular credible interval contains a parameter with measurable probability. Compare this to the confidence interval and its interpretation that, in the long run, a certain percentage of the realized confidence intervals will cover the true parameter. Many non-statisticians wrongly assume the Bayesian credible interval interpretation for a confidence interval interpretation.

The Bayesian approach also provides a way to build models and perform estimation and inference for complicated problems where using frequentist methods is cumbersome and sometimes not obvious. Hierarchical models and missing data problems are two cases that lend themselves to Bayesian solutions nicely. Although this paper is concerned with less sophisticated analyses in which the driving force is the desire for the Bayesian framework, it's important to note that the consummate value of the Bayesian method might be to provide statistical inference for problems that couldn't be handled without it.

Prior Distributions

Some practitioners want to benefit from the Bayesian framework with as limited an influence from the prior distribution as possible: this can be accomplished by choosing priors that have a minimal impact on the posterior distribution. Such priors are called noninformative priors, and they are popular for some applications, although they are not always easy to construct. An informative prior dominates the likelihood, and thus it has a discernible impact on the posterior distribution.

A prior distribution is noninformative if it is "flat" relative to the posterior distribution, as demonstrated in Figure 1. However, while a noninformative prior can appear to be more objective, it's important to realize that there is some degree of subjectivity in any prior chosen; it does not represent complete ignorance about the parameter in question. Also, using noninformative priors can lead to what is known as improper posteriors (nonintegrable posterior density), with which you cannot make inferences. Noninformative priors might also be noninvariant, which means that they could be noninformative in one parameterization but not noninformative if a transformation is applied.

On the other hand, an improper prior distribution, such as the uniform prior distribution on the number line, can be appropriate. Improper prior distributions are frequently used in Bayesian analysis because they yield noninformative priors and proper posterior distributions. To form a proper posterior distribution, the normalizing constant has to be finite for all $\mathbf{y}$.

Some of the priors available with the built-in Bayesian procedures are improper, but they all produce proper posterior distributions. However, the MCMC procedure enables you to construct whatever prior distribution you can program, and so you yourself have to ensure that the resulting posterior distribution is proper.

More about Priors

Jeffreys' prior (Jeffreys 1961) is a useful prior because it doesn't change much over the region in which the likelihood is significant and doesn't have large values outside that range—the local uniformity property. It is based on the observed Fisher information matrix. Because it is locally uniform, it is a noninformative prior. Thus, it provides an automated way of finding a noninformative prior for any parametric model; it is also invariant with respect to one-to-one transformations. The GENMOD procedure computes Jeffreys' prior for any generalized linear model, and you can use it for your prior density for any of the coefficient parameters. Jeffreys' prior can lead to improper posteriors, but not in the case of the PROC GENMOD usage.

You can show that Jeffreys' prior is

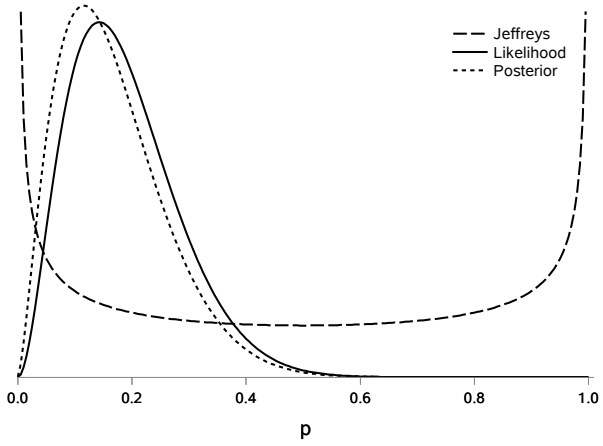
$$\pi(p) \propto p^{-1/2}(1-p)^{-1/2}$$

for the binomial distribution, and the posterior distribution for the vision example with Jeffreys' prior is

$$\begin{aligned} L(p)\pi(p) &\propto p^{y-\frac{1}{2}}(1-p)^{n-y-\frac{1}{2}} \\ &\sim \text{beta}(2.5, 12.5) \end{aligned}$$

Figure 3 displays how the Jeffreys' prior for the vision study example is relatively uninfluential in the areas where the posterior has the most mass.

Figure 3 Beta (2.5,12.5) Posterior with Jeffreys' Prior



A prior is a conjugate prior for a family of distributions if the prior and posterior distributions are from the same family. Conjugate priors result in closed-form solutions for the posterior distribution, enabling either direct inference or the construction of efficient Markov chain Monte Carlo sampling algorithms. Thus, the development of these priors was driven by the early need to minimize computational requirements. Although the computational barrier is no longer an issue, conjugate priors can still have performance benefits, and they are frequently used in Markov chain simulation because they directly sample from the target conditional distribution. For example, the GENMOD procedure uses conjugacy sampling wherever it is possible.

The beta is a conjugate prior for the binomial distribution. If the likelihood is based on the binomial (n, p):

$$L(p) \propto p^{y}(1-p)^{n-y}$$

and the prior is a beta (α, β),

$$\pi(p|\alpha, \beta) \propto p^{\alpha-1}(1-p)^{\beta-1}$$

then the posterior distribution is written as

$$\begin{aligned} \pi(p|\alpha, \beta, y, n) &\propto p^{y+\alpha-1}(1-p)^{n-y+\beta-1} \\ &= \text{beta}\,(y+\alpha, n-y+\beta) \end{aligned}$$

This posterior is easily calculated, and you can rely on simulations from it to produce the measures of interest, as demonstrated above.

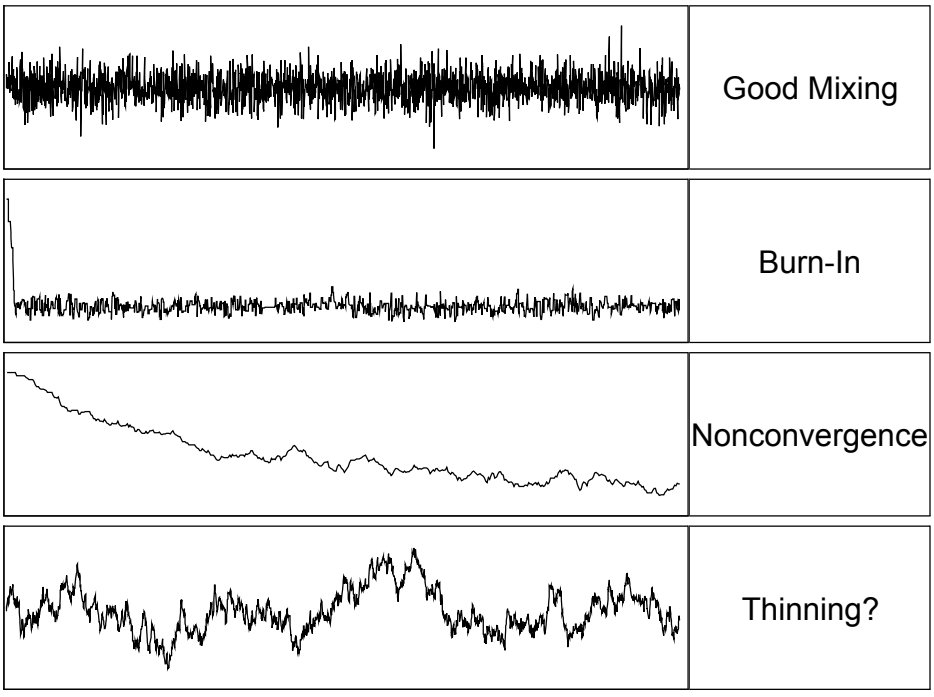
Assessing Convergence

Although this paper does not describe the underlying MCMC computations, and you can perform Bayesian analysis without knowing the specifics of those computations, it is important to understand that a Markov chain is being generated (its stationary distribution is the desired posterior distribution) and that you must check its convergence before you can work with the resulting posterior statistics. An unconverged Markov chain does not explore the parameter space sufficiently, and the samples cannot approximate the target distribution well. Inference should not be based on unconverged Markov chains, or misleading results can occur. And you need to check the convergence of all the parameters, not just the ones of interest.

There is no definitive way of determining that you have convergence, but there are a number of diagnostic tools that tell you if the chain hasn't converged. The built-in Bayesian procedures provide a number of convergence diagnostic tests and tools, such as Gelman-Rubin, Geweke, Heidelberger-Welch, and Raftery-Lewis tests. Autocorrelation measures the dependency among the Markov chain samples, and high correlations can indicate poor mixing. The Geweke statistic compares means from early and late parts of the Markov chain to see whether they have converged. The effective sample size (ESS) is particularly useful as it provides a numerical indication of mixing status. The closer ESS is to n, the better the mixing in the Markov chain. In general, an ESS of approximately 1,000 is adequate for estimating the posterior density. You might want it larger if you are estimating tail percentiles.

One of the ways that you can assess convergence is with visual examination of the trace plot, which is a plot of the sampled values of a parameter versus the sample number. Figure 4 displays some types of trace plots that can result:

Figure 4 Types of Trace Plots



By default, the built-in Bayesian procedures discard the first 2,000 samples as burn-in and keep the next 10,000. You want to discard the early samples to reduce the potential bias they might have on the estimates.(You can increase the number of samples when needed.) The first plot shows good mixing. The samples stay close to the high-density region of the target distribution; they move to the tail areas but

quickly return to the high-density region. The second plot shows evidence that a longer burn-in period is required. The third plot sets off warning signals. You could try increasing the number of samples, but sometimes a chain is simply not going to converge. Additional adjustment might be required such as model reparameterization or using a different sampling algorithm. Some practitioners might see thinning, or the practice of keeping every kth iteration to reduce autocorrelation, as indicated. However, current practice tends to downplay the usefulness of thinning in favor of keeping all the samples. The Bayesian procedures produce trace plots and also autocorrelation plots and density plots to aid in convergence assessment.

For further information about these measures, see the "Introduction to Bayesian Analysis" chapter in the *SAS/STAT User's Guide*.

Summary of the Steps in a Bayesian Analysis

You perform the following steps in a Bayesian analysis. Choosing a prior, checking convergence, and evaluating the sensitivity of your results to your prior might be new steps for many data analysts, but they are important ones.

1. Select a model (likelihood) and corresponding priors. If you have information about the parameters, use them to construct the priors.
2. Obtain estimates of the posterior distribution. You might want to start with a short Markov chain.
3. Carry out convergence assessment by using the trace plots and convergence tests. You usually iterate between this step and step 2 until you have convergence.
4. Check for the fit of the model and evaluate the sensitivity of your results due to the priors used.
5. Interpret the results: Do the posterior mean estimates make sense? How about the credible intervals?
6. Carry out further analysis: compare different models, or estimate various quantities of interest, such as functions of the parameters.

Bayesian Capabilities in SAS/STAT Software

SAS provides two avenues for Bayesian analysis: built-in Bayesian analysis in certain modeling procedures and the MCMC procedure for general-purpose modeling. The built-in Bayesian procedures are ideal for data analysts beginning to use Bayesian methods, and they suffice for many analysis objectives. Simply adding the BAYES statement generates Bayesian analyses without the need to program priors and likelihoods for the GENMOD, PHREG, LIFEREG, and FMM procedures. Thus, you can obtain Bayesian results for the following:

- linear regression
- Poisson regression
- logistic regression
- loglinear models
- accelerated failure time models
- Cox proportional models
- piecewise exponential models
- frailty models
- finite mixture models

The built-in Bayesian procedures apply the appropriate Markov chain Monte Carlo sampling technique. The Gamerman algorithm is the default sampling method for generalized linear models fit with the GENMOD procedure, and Gibbs sampling with adaptive rejection sampling (ARS) is generally the default, otherwise. However, conjugate sampling is available for a few cases, and the independent Metropolis algorithm and the random walk Metropolis algorithm are also available when appropriate.

The built-in Bayesian procedures provide default prior distributions depending on what models are specified. You can choose from other available priors by using the CPRIOR= option (for coefficient parameters) and SCALEPRIOR= option (for scale parameters). Other options allow you to choose the numbers of burn-ins, the number of iterations, and so on. The following posterior statistics are produced:

- point estimates: mean, standard deviation, percentiles
- interval estimates: equal-tail and highest posterior density (HPD) intervals
- posterior correlation matrix
- deviance information criteria (DIC)

All these procedures produce convergence diagnostic plots and statistics, and they are the same diagnostics that the MCMC procedure produces. You can also output posterior samples to a data set for further analysis. The following sections describe how to use the built-in Bayesian procedures to perform Bayesian analyses.

Linear Regression

Consider a study of 54 patients who undergo a certain type of liver operation in a surgical unit (Neter el al 1996). Researchers are interested in whether blood clotting score has a positive effect on survival.

The following statements create SAS data set SURGERY. The variable Y is the survival time, and LOGX1 is the natural logarithm of the blood clotting score.

```
data surgery;
   input x1 logy;
   y = 10**logy;
   label x1 = 'Blood Clotting Score';
   label y = 'Survival Time';
   logx1 = log(x1);
datalines;
6.7   2.3010
5.1   2.0043
..
..
;
run;
```

Suppose you want to perform a Bayesian analysis for the following regression model for the survival times, where ϵ is a $N(0, \sigma^2)$ error term:

$$Y = \beta_0 + \beta_1 logX_1 + \epsilon$$

If you wanted a frequentist analysis, you could fit this model by using the REG procedure. But this model is also a generalized linear model (GLM) with a normal distribution and the identity link function, so it can be fit with the GENMOD procedure, which offers Bayesian analysis. To review, a GLM relates a mean response to a vector of explanatory variables through a monotone link function where the likelihood function belongs to the exponential family. The link function g describes how the expected value of the response variable is related to the linear predictor,

$$g(E(y_i)) = g(\mu_i) = x_i^t \beta$$

where y_i is a response variable ($i = 1, \ldots, n$), g is a link function, $\mu_i = E(y_i)$, x_i is a vector of independent variables, and β is a vector of regression coefficients to be estimated. For example, when you assume a normal distribution for the response variable, you specify an identity link function $g(\mu) = \mu$. For Poisson regression you specify a log link function $g(\mu) = \log(\mu)$, and for a logistic regression you specify a logit link function $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$.

The BAYES statement produces Bayesian analyses with the GENMOD procedure for most of the models it fits; currently this does not include models for which the multinomial distribution is used. The first step of a Bayesian analysis is specifying the prior distribution, and Table 2 describes priors that are supported by the GENMOD procedure.

Table 2 Prior Distributions Provided by the GENMOD Procedure

Parameter	Prior
Regression coefficients	Jeffreys', normal, uniform
Dispersion	Gamma, inverse gamma, improper
Scale and precision	Gamma, improper

You would specify a prior for one of the dispersion, scale, or precision parameters, in models that have such parameters.

The following statements request a Bayesian analysis for the linear regression model with PROC GENMOD:

```
proc genmod data=surg;
   model y=logx1/dist=normal;
   bayes seed=1234 outpost=Post;
run;
```

The MODEL statement with the DIST=NORMAL option describes the simple linear regression model (the default is the identity link function.) The BAYES statement requests the Bayesian analysis. The SEED= option in the BAYES statement sets the random number seed so you can reproduce the analysis in the future. The OUTPOST= option saves the generated posterior samples to the POST data set for further analysis.

By default, PROC GENMOD produces the maximum likelihood estimates of the model parameters, as displayed in Figure 5.

Figure 5 Maximum Likelihood Parameter Estimates

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept	1	-94.9822	114.5279	-319.453	129.4884
logx1	1	170.1749	65.8373	41.1361	299.2137
Scale	1	135.7963	13.0670	112.4556	163.9815

Note: The scale parameter was estimated by maximum likelihood.

Subsequent tables are produced by the Bayesian analysis. The "Model Information" table in Figure 6 summarizes information about the model that you fit. The 2,000 burn-in samples are followed by 10,000 samples. Because the normal distribution was specified in the MODEL statement, PROC GENMOD exploited the conjugacy and sampled directly from the target distribution.

Figure 6 Model Information

Bayesian Analysis

Model Information		
Data Set	WORK.SURGERY	
Burn-In Size	2000	
MC Sample Size	10000	
Thinning	1	
Sampling Algorithm	Conjugate	
Distribution	Normal	
Link Function	Identity	
Dependent Variable	y	Survival Time

The "Prior Distributions" table in Figure 7 identifies the prior distributions. The default uniform prior distribution is assumed for the regression coefficients β_0 and β_1, and the default improper prior is used for the dispersion parameter. An improper prior is defined as

$$\pi(\theta) \propto \frac{1}{\theta}$$

Both of these priors are noninformative.

Figure 7 Prior Distributions

Bayesian Analysis

Uniform Prior for Regression Coefficients

Parameter	Prior
Intercept	Constant
logx1	Constant

Independent Prior Distributions for Model Parameters

Parameter	Prior Distribution
Dispersion	Improper

Figure 8 displays the convergence diagnostics: the "Posterior Autocorrelations" table reports that the autocorrelations at the selected lags (1, 5, 10, and 50, by default) drop off quickly, indicating reasonable mixing of the Markov chain. The p-values in the "Geweke Diagnostics" table show that the mean estimate of the Markov chain is stable over time. The "Effective Sample Sizes" table reports that the number of effective sample sizes of the parameters is equal to the Markov chain sample size, which is as good as that measure gets.

Figure 8 Convergence Diagnostics Table

Bayesian Analysis

Posterior Autocorrelations

Parameter	Lag 1	Lag 5	Lag 10	Lag 50
Intercept	0.0059	0.0145	-0.0059	0.0106
logx1	0.0027	0.0124	-0.0046	0.0091
Dispersion	0.0002	-0.0031	0.0074	0.0014

Geweke Diagnostics

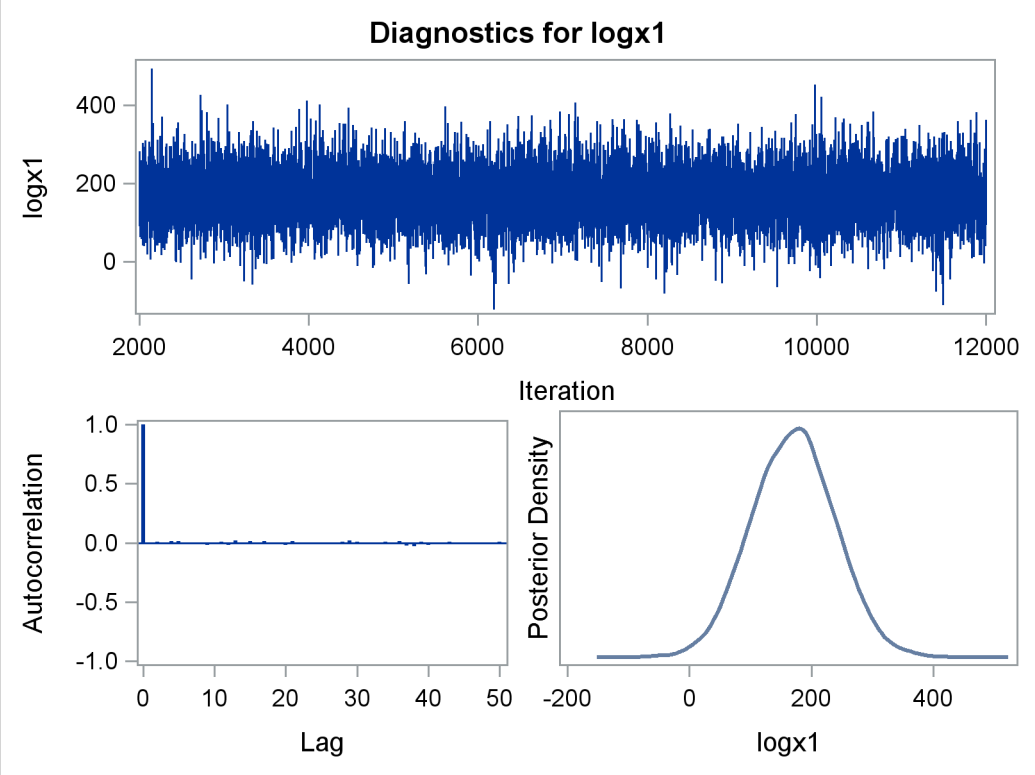
Parameter	z	Pr > |z|
Intercept	-0.2959	0.7673
logx1	0.2873	0.7739
Dispersion	0.9446	0.3449

Effective Sample Sizes

Parameter	ESS	Autocorrelation Time	Efficiency
Intercept	10000.0	1.0000	1.0000
logx1	10000.0	1.0000	1.0000
Dispersion	10000.0	1.0000	1.0000

The built-in Bayesian procedures produce three types of plots that help you visualize the posterior samples of each parameter. These diagnostics plots for the slope coefficient β_1 are shown in Figure 9. The trace plot indicates that the Markov chain has stabilized with good mixing. The autocorrelation plot confirms the tabular information, and the kernel density plot estimates the posterior marginal distribution. The diagnostic plots for the other parameters (not shown here) have similar outcomes.

Figure 9 Bayesian Model Diagnostic Plot for β_1



Because convergence doesn't seem to be an issue, you can review the posterior statistics displayed in Figure 10 and Figure 11.

Figure 10 Posterior Summary Statistics

Bayesian Analysis

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
Intercept	10000	-95.9018	119.4	-176.1	-96.0173	-16.2076
logx1	10000	170.7	68.6094	124.2	171.1	216.7
Dispersion	10000	19919.9	4072.4	17006.0	19421.7	22253.7

Figure 11 Posterior Interval Statistics

Posterior Intervals

Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Intercept	0.050	-328.7	135.5	-324.0	139.2
logx1	0.050	37.4137	304.3	36.8799	303.0
Dispersion	0.050	13475.4	29325.9	12598.8	28090.3

The posterior summaries displayed in Figure 10 are similar to the maximum likelihood estimates shown in Figure 5. This is because noninformative priors were used and the posterior is effectively the likelihood. Figure 11 displays the HPD interval and the equal-tail interval.

You might be interested in whether LOGX1 has a positive effect on survival time. You can address this question by using the posterior samples that are saved to the POST data set. This means that you can determine the conditional probability $\Pr(\beta_1 > 0 \mid \mathbf{y})$ directly from the posterior sample. All you have to do is to determine the proportions of samples where $\beta_1 > 0$,

$$\Pr(\beta_1 > 0 \mid \mathbf{y}) = \frac{1}{N} \sum_{t=1}^{N} I(\beta_1^t > 0)$$

where N = 10000 is the number of samples after burn-in and I is an indicator function, where $I(\beta_1^t > 0) = 1$ if $\beta_1^t > 0$, and 0 otherwise. The following SAS statements produce the posterior samples of the indicator function I by using the posterior samples of β_1 saved in the output data set POST:

```
data Prob;
   set Post;
   Indicator = (logX1 > 0);
   label Indicator= 'log(Blood Clotting Score) > 0';
run;
```

The following statements request the summary statistics by using the MEANS procedure:

```
proc means data = prob(keep=Indicator);
run;
```

Figure 12 displays the results. The posterior probability that β_1 greater than 0 is estimated as 0.9936. Obviously LOGX1 has a strongly positive effect on survival time.

Figure 12 Posterior Summary Statistics with the MEANS Procedure

Analysis Variable : Indicator log(Blood Clotting Score) > 0				
N	Mean	Std Dev	Minimum	Maximum
10000	0.9936000	0.0797476	0	1.0000000

Logistic Regression

Consider a study of the analgesic effects of treatments on elderly patients with neuralgia. A test treatment and a placebo are compared, and the response is whether the patient reports pain. Explanatory variables include the age and gender of the 60 patients and the duration of the complaint before the treatment began.

The following SAS statements input the data:

```
data Neuralgia;
   input Treatment $ Sex $ Age Duration Pain $ @@;
datalines;
P  F  68   1  No   B  M  74  16  No  P  F  67  30  No
P  M  66  26  Yes  B  F  67  28  No  B  F  77  16  No
A  F  71  12  No   B  F  72  50  No  B  F  76   9  Yes
A  M  71  17  Yes  A  F  63  27  No  A  F  69  18  Yes
..
..
;
```

Logistic regression is considered for this data set:

$$\begin{aligned} \text{pain}_i &\sim \text{binary}(p_i) \\ p_i &= \text{logit}(\beta_0 + \beta_1 \cdot \text{Sex}_{F,i} + \beta_2 \cdot \text{Treatment}_{A,i} \\ &\quad + \beta_3 \cdot \text{Treatment}_{B,i} + \beta_4 \cdot \text{Sex}_{F,i} \cdot \text{Treatment}_{A,i} \\ &\quad + \beta_5 \cdot \text{Sex}_{F,i} \cdot \text{Treatment}_{B,i} + \beta_6 \cdot \text{Age} + \beta_7 \cdot \text{Duration}) \end{aligned}$$

where Sex_F, Treatment_A, and Treatment_B are dummy variables for the categorical predictors.

You might consider a normal prior with large variance as a noninformative prior distribution on all the regression coefficients:

$$\pi(\beta_0, \cdots, \beta_7) \sim \text{normal}(0, \text{var} = 1e6)$$

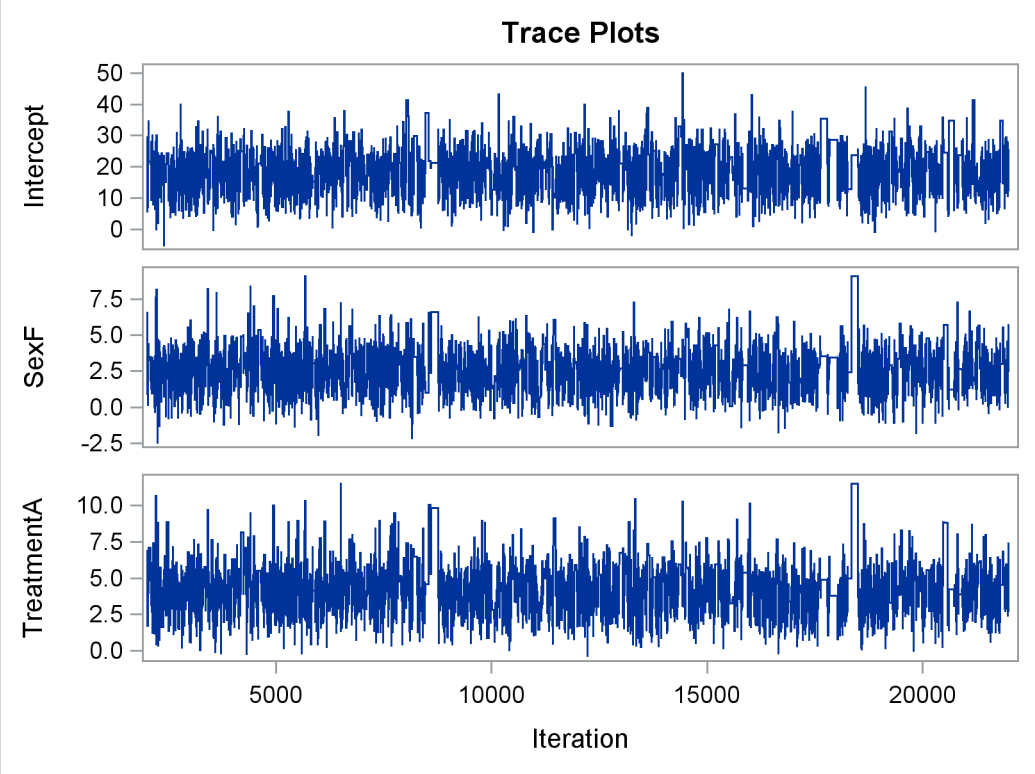
You can also fit this model with the GENMOD procedure. The following statements specify the analysis:

```
proc genmod data=neuralgia;
   class Treatment(ref="P") Sex(ref="M");
   model Pain= sex|treatment Age Duration / dist=bin link=logit;
   bayes seed=1 cprior=normal(var=1e6) outpost=neuout plots=trace nmc=20000;
run;
```

The CPRIOR=NORMAL(VAR=1E6) option specifies the normal prior for the coefficients; the specified large variance is requested with VAR=1E6. The PLOTS=TRACE option requests only the trace plots. Logistic regression is requested with the DIST=BIN and the LINK=LOGIT options (either one will do). The default sampling algorithm for generalized linear models is the Gamerman algorithm, which uses an iterative weighted least squares algorithm to sample the coefficients from their conditional distributions. It usually performs very well. The NMC option requests 20,000 samples and is specified because the default number of simulations results in low ESS.

Figure 13 displays the trace plots for some of the parameters. They show good mixing.

Figure 13 Trace Plots



The effective sample sizes are adequate and do not indicate any issues in the convergence of the Markov chain, as seen in Figure 14. With 20,000 samples, the chain has stabilized appropriately.

Figure 14 Effective Sample Sizes

Effective Sample Sizes

Parameter	ESS	Autocorrelation Time	Efficiency
Intercept	685.2	29.1890	0.0343
SexF	569.3	35.1325	0.0285
TreatmentA	401.8	49.7762	0.0201
TreatmentB	491.7	40.6781	0.0246
TreatmentASexF	596.4	33.5373	0.0298
TreatmentBSexF	585.2	34.1753	0.0293
Age	604.3	33.0943	0.0302
Duration	1054.7	18.9619	0.0527

Thus, the posterior summaries are of interest. Figure 15 displays the posterior summaries and Figure 16 displays the posterior intervals.

Figure 15 Posterior Summaries

Bayesian Analysis

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
Intercept	20000	19.6387	7.5273	14.2585	19.6414	24.4465
SexF	20000	2.7964	1.6907	1.6403	2.6644	3.7348
TreatmentA	20000	4.5857	1.9187	3.3402	4.4201	5.5466
TreatmentB	20000	5.0661	1.9526	3.7114	4.8921	6.2388
TreatmentASexF	20000	-1.0052	2.3403	-2.3789	-0.9513	0.5297
TreatmentBSexF	20000	-0.2559	2.2854	-1.7394	-0.2499	1.2874
Age	20000	-0.3365	0.1134	-0.4192	-0.3316	-0.2522
Duration	20000	0.00790	0.0375	-0.0164	0.00668	0.0312

Figure 16 Posterior Intervals

Posterior Intervals

Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Intercept	0.050	5.7037	35.4140	4.0078	32.9882
SexF	0.050	-0.0605	6.5737	0.0594	6.6094
TreatmentA	0.050	1.3776	9.4931	1.2502	8.9476
TreatmentB	0.050	1.6392	9.2298	1.7984	9.3168
TreatmentASexF	0.050	-5.8694	3.4826	-5.9435	3.2389
TreatmentBSexF	0.050	-4.7975	4.3188	-4.1485	4.4298
Age	0.050	-0.5721	-0.1338	-0.5422	-0.1159
Duration	0.050	-0.0637	0.0869	-0.0669	0.0821

The intervals suggest that treatment is highly influential, and although age appears to be important, other covariates and the sex $\times$ treatment interactions do not appear to be important. However, all terms are kept in the model.

Odds ratios are an important measure of association that can be estimated by forming functions of the parameter estimates for a logistic regression. See Stokes, Davis, and Koch (2012) for a full discussion. For example, if you want to form an odds ratio comparing the odds of no pain for the female patients to the odds of no pain for the male patients, you can compute it by exponentiating the corresponding linear combination of the relevant model parameters.

Because a function of random variables is also a random variable, you can compute odds ratios from the results of a Bayesian analysis by manipulating the posterior samples. You save the results of your analysis to a special data set known as a SAS item store using the STORE statement, and then you compute the odds ratio by using the ESTIMATE statement of the PLM procedure. PROC PLM performs postfitting tasks by operating on the posterior samples from PROC GENMOD. (For an MLE analysis, it operates on the saved parameter estimates and covariances.) Recall that the ESTIMATE statement provides custom hypotheses, which means that it can also be used to form functions from the linear combinations of the regression coefficients.

The following statements fit the same model with PROC GENMOD and saves the posterior samples to the LOGIT_BAYES item store:

```
proc genmod data=neuralgia;
   class Treatment(ref="P") Sex(ref="M");
   model Pain= sex|treatment Age Duration / dist=bin link=logit;
   bayes seed=2 cprior=normal(var=1e6) outpost=neuout
```

```
   plots=trace algorithm=im;
   store logit_bayes;
run;
```

Note that a different sampler is requested with the SAMPLING=IM option, which requests the Independence Metropolis sampler, which is a variation of the Metropolis sampling algorithm. This sampler finds an envelope distribution to the posterior and uses it as an efficient proposal distribution to generate the posterior samples. It is generally faster than the Gamerman algorithm, but the latter is more computationally stable so it remains the default sampler in PROC GENMOD. In this case, the IM sampler produces better ESS values.

The next set of statements inputs the LOGIT_BAYES item store into the PLM procedure with the RESTORE= option. The desired odds ratio estimate is specified with the ESTIMATE statement. Because this model includes the SEX*TREATMENT interaction, the odds ratio requested is the one that compares female and male odds at the A treatment level. The 1 and -1 coefficients compare females and males for the overall gender effect and also at the A level for treatment. You are effectively creating a distribution of these functions from each of the posterior sample estimates. Plots are provided by the PLM procedure when it is executed in this sample-based mode, and the EXP option requests the exponentiation that produces the odds ratio estimate from the parameters. The CL option requests 95% credible intervals.

```
proc plm restore=logit_bayes;
   estimate "F vs M, at Trt=A"
     sex 1 -1 treatment*sex 1 -1 0 0 0 0
   / exp cl;
run;
```

Figure 17 displays the coefficients of the L matrix for confirmation that the appropriate contrast was specified.

Figure 17 L Matrix Coefficients

Estimate Coefficients			
Parameter	**Treatment**	**Sex**	**Row1**
Intercept			
Sex F		F	1
Sex M		M	-1
Treatment A	A		
Treatment B	B		
Treatment P	P		
Treatment A * Sex F	A	F	1
Treatment A * Sex M	A	M	-1
Treatment B * Sex F	B	F	
Treatment B * Sex M	B	M	
Treatment P * Sex F	P	F	
Treatment P * Sex M	P	M	
Age			
Duration			

Figure 18 displays the odd ratio estimate in the “Exponentiated” column, which has the value 28.39. This says that the odds of no pain are 28 times higher for females than for males at the A treatment level.

Figure 18 Posterior Odds Ratio Estimate

Label	N	Sample Estimate Estimate	Standard Deviation	Percentiles 25th	50th	75th	Alpha	Lower HPD	Upper HPD	Exponentiated	Standard Deviation of Exponentiated
F vs M, at Trt=A	10000	1.8781	1.5260	0.7768	1.7862	2.9174	0.05	-0.7442	4.9791	28.3873	188.824003

Label	Sample Estimate Percentiles for Exponentiated 25th	50th	75th	Lower HPD of Exponentiated	Upper HPD of Exponentiated
F vs M, at Trt=A	2.1744	5.9664	18.4925	0.1876	93.1034

Proportional Hazards Model

Consider the data from the Veteran Adminstration's lung cancer trial that are presented in Appendix 1 of Kalbfleisch and Prentice (1980). Death in days is the response, and type of therapy, type of tumor cell, presence of prior therapy, age, and time from diagnosis to study inclusion are included in the data set. There is also an indicator variable for whether the response was censored.

The first 20 observations of SAS data set VALUNG are shown in Figure 19.

below:

Figure 19 Subset of VALUNG Data

Obs	Therapy	Cell	Time	Kps	Duration	Age	Ptherapy	Status
1	standard	squamous	72	60	7	69	no	1
2	standard	squamous	411	70	5	64	yes	1
3	standard	squamous	228	60	3	38	no	1
4	standard	squamous	126	60	9	63	yes	1
5	standard	squamous	118	70	11	65	yes	1
6	standard	squamous	10	20	5	49	no	1
7	standard	squamous	82	40	10	69	yes	1
8	standard	squamous	110	80	29	68	no	1
9	standard	squamous	314	50	18	43	no	1
10	standard	squamous	100	70	6	70	no	0
11	standard	squamous	42	60	4	81	no	1
12	standard	squamous	8	40	58	63	yes	1
13	standard	squamous	144	30	4	63	no	1
14	standard	squamous	25	80	9	52	yes	0
15	standard	squamous	11	70	11	48	yes	1
16	standard	small	30	60	3	61	no	1
17	standard	small	384	60	9	42	no	1
18	standard	small	4	40	2	35	no	1
19	standard	small	54	80	4	63	yes	1
20	standard	small	13	60	4	56	no	1

Interest lies in how therapy and the other covariates impact time until death, and the analysis is further complicated because the response is censored. The proportional hazards model, or Cox regression model, is widely used in the analysis of time-to-event data to explain the effect of explanatory variables on hazard rates. The survival time of each member of a population is assumed to follow its own hazard function $\lambda_i(t)$,

$$\lambda_l(t) = \lambda(t; Z_l) = \lambda_0(t)\exp(Z_l'\beta)$$

where $\lambda_0(t)$ is an arbitrary and unspecified baseline hazard function, Z_l is the vector of explanatory variables for the lth individual, and β is the vector of unknown regression parameters (which is assumed to be the same for all individuals). The PHREG procedure fits the Cox model by maximizing the partial likelihood function; this eliminates the unknown baseline hazard $\lambda_0(t)$ and accounts for censored survival times. In the Bayesian approach, the partial likelihood function is used as the likelihood function in the posterior distribution (Sinha, Ibrahim, and Chen 2003).

The PHREG procedure supports the following priors for proportional hazards regression, displayed in Table 3.

Table 3 Prior Distributions for the Cox Model

Parameter	Prior
Regression coefficients	Normal, uniform, Zellner's g-prior
Zellner's g	Constant, gamma

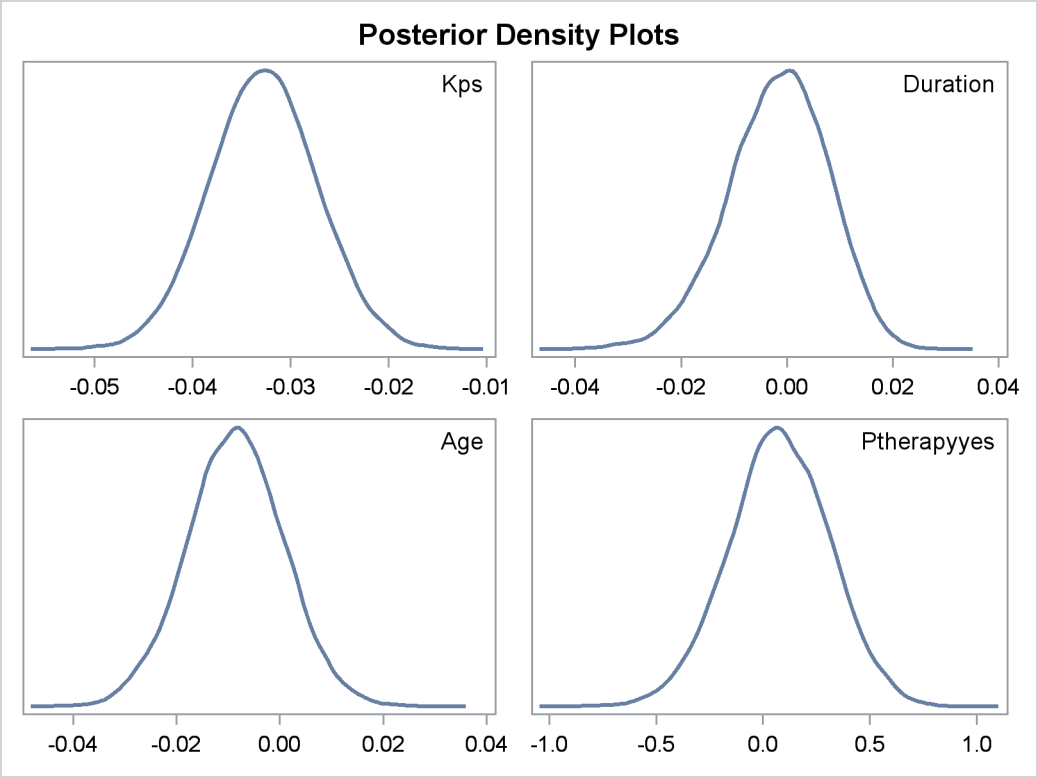
The following PROC PHREG statements request Bayesian analysis for the Cox proportional hazard model with uniform priors on the regression coefficients. The categorical variables PTHERAPY, CELL, and THERAPY are declared in the CLASS statement. The value "large" is specified as the reference category for type of tumor cell, the value "standard" is specified as the reference category for therapy, and the value "no" is the reference category for prior therapy.

```
   proc phreg data=VALung;
    class Ptherapy(ref='no') Cell(ref='large') Therapy(ref='standard');
    model Time*Status(0) = Kps Duration Age Ptherapy Cell Therapy;
    hazardratio 'HR' Therapy;
    bayes seed=1 outpost=out cprior=uniform plots=density;
run;
```

The BAYES statement requests Bayesian analysis. The CPRIOR= option specifies the uniform prior for the regression coefficients, and the PLOTS= option requests a panel of posterior density plots for the model parameters. The HAZARDRATIO statement requests the hazard ratio comparing the odds for therapies.

Figure 20 shows kernel posterior density plots that estimate the posterior marginal distributions for the first four regression coefficients. Each plot shows a smooth, unimodal shape for the posterior marginal distribution.

Figure 20 Density Plots



Posterior summary statistics and intervals are shown in Figure 21. Because the prior distribution is noninformative, the results are similar to those obtained with the standard PROC PHREG analysis (not shown here).

Figure 21 Posterior Summary Statistics

Bayesian Analysis

Posterior Summaries and Intervals

Parameter	N	Mean	Standard Deviation	95% HPD Interval	
Kps	10000	-0.0327	0.00545	-0.0434	-0.0221
Duration	10000	-0.00170	0.00945	-0.0202	0.0164
Age	10000	-0.00852	0.00935	-0.0270	0.00983
Ptherapyyes	10000	0.0754	0.2345	-0.3715	0.5488
Celladeno	10000	0.7867	0.3080	0.1579	1.3587
Cellsmall	10000	0.4632	0.2731	-0.0530	1.0118
Cellsquamous	10000	-0.4022	0.2843	-0.9550	0.1582
Therapytest	10000	0.2897	0.2091	-0.1144	0.6987

The posterior mean of the hazard ratio in Figure 22 indicates that the standard treatment is more effective than the test treatment for individuals who have the given covariates. However, note that both the 95% equal-tailed interval and the 95% HPD interval contain the hazard ratio value of 1, which suggests that the test therapy is not really different from the standard therapy.

Figure 22 Hazard Ratio and Confidence Limits

HR: Hazard Ratios for Therapy

				Quantiles						
Description	N	Mean	Standard Deviation	25%	50%	75%	95% Equal-Tail Interval		95% HPD Interval	
Therapy standard vs test	10000	0.7651	0.1617	0.6509	0.7483	0.8607	0.4988	1.1265	0.4692	1.0859

Suppose you are interested in estimating the survival curves for two individuals who have similar characteristics, one of whom receives the test treatment while the other receives the standard treatment. The following SAS DATA step creates the PRED data set which includes the covariate values for these individuals:

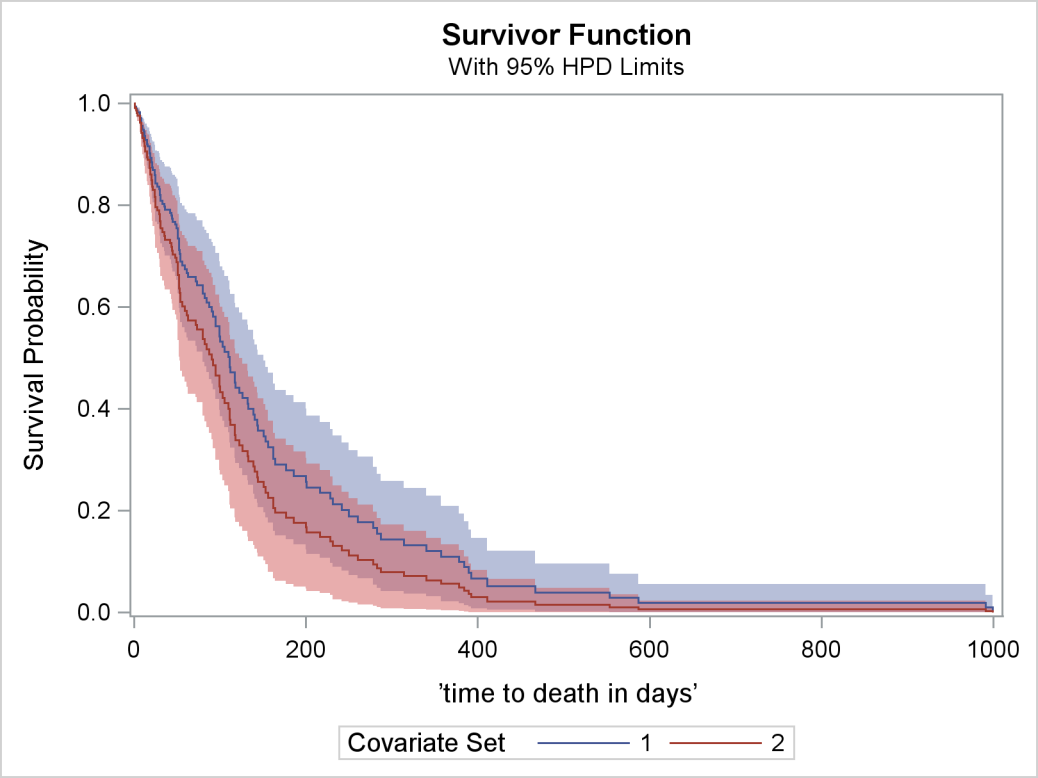
```
data Pred;
   input Ptherapy  Kps Duration Age Cell $ Therapy $ @@;
   format Ptherapy yesno.;
   datalines;
   0 58 8.7 60 large standard
   0 58 8.7 60 large test
   ;
```

The following statements request the estimation of separate survival curves for these two sets of covariates. The PLOTS= option in the PROC PHREG statement requests survival curves, the OVERLAY suboption overlays the curves in the same plot, and the CL= suboption requests the highest posterior density (HPD) confidence limits for the survivor functions. The COVARIATES= option in the BASELINE statement specifies the covariates data set PRED.

```
proc phreg data=VALung plots(overlay cl=hpd)=survival;
   baseline covariates=Pred;
   class Therapy(ref='standard') Cell(ref='large') Ptherapy(ref='no');
   model Time*Status(0) = Kps Duration Age Ptherapy Cell Therapy;
   bayes seed=1 outpost=out cprior=uniform plots=density;
run;
```

Figure 23 displays the survival curves for these sets of covariate values along with their HPD confidence limits.

Figure 23 Posterior Survival Curves



Capabilities of SAS Built-In Bayesian Procedures

The examples in this paper provide an introduction to the capabilities of the built-in Bayesian procedures in SAS/STAT. But not only can the GENMOD procedure produce Bayesian analyses for linear regression and logistic regression, it also provides Bayesian analysis for Poisson regression, negative binomial regression, and models based on the Tweedie distribution. While PROC GENMOD doesn't offer Bayesian analysis for zero-inflated models, such as the zero-inflated Poisson model, you can still use the built-in Bayesian procedures by switching to the FMM procedure, which fits finite mixture models, which includes zero-inflated models.

The FMM procedure provides Bayesian analysis for finite mixtures of normal, T, binomial, binary, Poisson, and exponential distributions. This covers ground from zero-inflated models to clustering for univariate data. PROC FMM applies specialized sampling algorithms, depending on the distribution family and the nature of the model effects, including conjugate samplers and the Metropolis-Hastings algorithm. PROC FMM provides a single parametric form for the prior but you can adjust it to reflect specific prior information. It provides posterior summaries and convergence diagnostics similar to those produced by the other built-in Bayesian procedures. See Kessler and McDowell (2012) for an example of Bayesian analysis using the FMM procedure.

This paper illustrates the use of the PHREG procedure to perform Bayesian analysis for Cox regression; it can handle time-dependent covariates and all TIES= methods. (The Bayesian functionality does not currently apply to models that have certain data constraints such as recurrent events.) Bayesian analysis is also available for the frailty models fit with the PHREG procedure, which enable you to analyze survival times that are clustered. In addition, PROC PHREG also provides Bayesian analysis for the piecewise exponential model.

The LIFEREG procedure provides Bayesian analysis for parametric lifetime models. This capability is now available for the exponential, 3-parameter gamma, log-logistic, log-normal, logistic, normal, and Weibull distributions. Model parameters are the regression coefficients and dispersion parameters (or precision or scale). Normal and uniform priors are provided for the regression coefficients, and the gamma and improper priors are provided for the scale and shape parameters.

General Bayesian Modeling with the MCMC Procedure

Although the built-in Bayesian procedures provide Bayesian analyses for many standard techniques, they only go so far. You might want to include priors that are not offered, or you might want to perform a Bayesian analysis for a model that isn't covered. For example, the GENMOD procedure doesn't presently offer Bayesian analysis for the proportional odds model. However, you can fit nearly any model that you want, for any prior and likelihood you can program, with the MCMC procedure.

The MCMC procedure is a general-purpose Bayesian modeling tool. It was built on the familiar syntax of programming statements used in the NLMIXED procedure. PROC MCMC performs posterior sampling and statistical inference for Bayesian parametric models. The procedure fits single-level or multilevel models. These models can take various forms, from linear to nonlinear models, by using standard or nonstandard distributions. Using PROC MCMC requires using programming statement to declare the parameters in the model, to specify prior distributions for them, and to describe the conditional distribution for the response variable given the parameters. You can specify a model by using keywords for a standard form (normal, binomial, gamma) or use programming statements to define a general distribution.

The MCMC procedure uses various algorithms to simulate samples from the model that you specify. You can also choose an optimization technique (such as the quasi-Newton algorithm) to estimate the posterior mode and approximate the covariance matrix around the mode. PROC MCMC computes a number of posterior estimates, and it also outputs the posterior samples to a data set for further analysis. The MCMC procedure provides the same convergence diagnostics that are produced by using a BAYES statement in the built-in Bayesian procedures, as discussed previously.

The RANDOM statement in the MCMC procedure facilitates the specification of random effects in hierarchical linear or nonlinear models. You can build nested or nonnested hierarchical models to arbitrary depth. Using the RANDOM statement can result in reduced simulation time and improved convergence for models that have a large number of subjects. The MCMC procedure also handles missing data for both responses and covariates. Chen (2009, 2011, 2013) provides good overviews of the capabilities of the MCMC procedure, which continues to be enhanced. The most recent release of PROC MCMC in SAS/STAT 13.1 is multithreaded, providing faster performance for many models.

The BCHOICE Procedure

The first SAS/STAT procedure designed to focus on Bayesian analysis for a specific application is the BCHOICE procedure. PROC BCHOICE performs Bayesian analysis for discrete choice models. Discrete choice models are used in marketing research to model decision makers' choices among alternative products and services. The decision makers might be people, households, companies, and so on, and the alternatives might be products, services, actions, or any other options or items about which choices must be made. The collection of alternatives is known as a choice set, and when individuals are asked to choose, they usually assign a utility to each alternative. The BCHOICE procedure provides Bayesian discrete choice models such as the multinomial logit, multinomial logit with random effects, and the nested logit. The probit response function is also available. You can supply a prior distribution for the parameters if you want something other than the default noninformative prior. PROC BCHOICE obtains samples from the corresponding posterior distributions, produces summary and diagnostic statistics, and saves the posterior samples to an output data set that can be used for further analysis. PROC BCHOICE is also multithreaded.

Summary

The SAS built-in Bayesian procedures provide a great deal of coverage for standard statistical analyses. With a ready set of priors and carefully-chosen default samplers, they make Bayesian computing very convenient for the SAS/STAT user. They provide a great starting place for the statistician who needs some essential Bayesian analysis now and might want to add more sophisticated Bayesian computing skills later. Bayesian analysis is an active area in SAS statistical software development, and each new release brings additional features.

For more information, the "Introduction to Bayesian Analysis" chapter in the *SAS/STAT User's Guide* contains a reading list with comprehensive references, including many references at the introductory level. In addition, the Statistics and Operations Focus Area includes substantial information about the statistical products, and

you can find it at support.sas.com/statistics/. The quarterly e-newsletter for that site is available on its home page. And of course, complete information is available in the online documentation located at support.sas.com/documentation/onlinedoc/stat/. The features of each new release are described in a "What's New" chapter that is must reading for SAS/STAT users.

References

Chen, F. (2009), "Bayesian Modeling Using the MCMC Procedure," in *Proceedings of the SAS Global Forum 2008 Conference,* Cary NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings09/257-2009.pdf.

Chen, F. (2011), "The RANDOM Statement and More: Moving on with PROC MCMC," in *Proceedings of the SAS Global Forum 2011 Conference,* Cary NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings11/334-2011.pdf.

Chen, F. (2013), "Missing No More: Using the MCMC Procedure to Model Missing Data", in *Proceedings of the SAS Global Forum 2013 Conference,* Cary NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings11/436-2013.pdf.

Gamerman, D. (1997), "Sampling from the Posterior Distribution in Generalized Linear Mixed Models,"*Statistics and Computing*, 7, 57–68.

Gilks, W. R. and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995), "Adaptive Rejection Metropolis Sampling," *Applied Statistics*, 44, 455–472.

Kalbfleisch, J. D. and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.

Kessler, D. and McDowell, A. (2012), "Introducing the FMM Procedure for Finite Mixture Models," in *Proceedings of the SAS Global Forum 2012 Conference,* Cary NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings12/328-2012.pdf.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996), *Applied Linear Statistical Models,* Fourth Edition, Chicago: Irwin.

Sinha, D., Ibrahim, J. G., and Chen, M. (2003), "Bayesian Justification of Cox's Partial Likelihood", *Biometrics*, 90, 629–641.

Stokes, M. E., Davis, C. E., and Koch G G. (2012). *Categorical Data Analysis Using SAS, Third Edition*, SAS Press: Cary NC.

Acknowledgments

The authors are grateful to Ed Huddleston for editing this manuscript. They would also like to thank David Kessler and Warren Kuhfeld for their helpful suggestions.

Contact Information

Your comments and questions are valued and encouraged. Contact the author:

Maura Stokes
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513

Version

1.0