Introduction to Bayesian Analysis Using SAS® Software

Joseph G. Ibrahim
Department of Biostatistics
University of North Carolina
Outline

1 Introduction to Bayesian statistics
   - Background and concepts in Bayesian methods
   - Prior distributions
   - Advanced computational methods
   - Convergence diagnostics and summary statistics
   - Discussion and summary

2 The GENMOD, LIFEREG, and PHREG procedures
Why Bayesian, Why Now?

Philosophical differences aside, there are practical reasons that drive the recent popularity of Bayesian analysis.

- Simplicity in thinking about problems and answering scientific questions
- Flexibility in making inference on a wide range of models, such as data augmentation and hierarchical models
- Incorporation of prior information
- Heavy promotion by both practitioners and theoretical statisticians
- Development of efficient inference and sampling tools
- Fast computers
Historical Highlights

- Bayes theorem was developed by the Reverend Thomas Bayes (1702–1761). His paper was published in 1763, posthumously.
- He formulated a way to solve the question of inverse probability: after observing a collection of Events, what is Pr(one Event)?
- Given a probability distribution of $p(x|\theta)$, where $x$ is an observable quantity, the inverse probability is the probability distribution of an unobserved variable $\theta$ given $x$, $p(\theta|x)$.
- The term “Bayesian” was coined around 1950, by R. A. Fisher.
Two Different Points of View\textsuperscript{1}

**Frequentist/Classical**

- Probabilities are objective properties of the real world. Probability refers to limiting relative frequencies.
- Parameters $\theta$ are fixed, unknown constants.
- Statistical procedures should be designed to have well-defined long-run frequency properties, such as the confidence interval.

\textsuperscript{1}Wasserman 2004
Two Different Points of View

Bayesian

- Probability describes degree of belief, not limiting frequency. The probability of an event is the degree to which you believe that the event is true. It is subjective.
- Parameters cannot be determined exactly. They are random variables, and you can make probability statements about them.
- Inferences about $\theta$ are based on the probability distribution for the parameter.
The Bayesian Method

Given data \( \mathbf{x} = \{x_1, \ldots, x_n\} \), Bayesian inference is carried out in the following way:

1. You choose a **prior distribution** \( \pi(\theta) \) for \( \theta \). The distribution describes your beliefs about the parameter *before* you examine the data.

2. Given the observed data \( \mathbf{x} \), you select a model (density) \( f(\mathbf{x}|\theta) \) to describe the distribution of \( \mathbf{x} \) given \( \theta \). The **Likelihood function** is any function proportional to \( f(\mathbf{x}|\theta) \); that is, \( \mathcal{L}(\theta) \propto f(\mathbf{x}|\theta) \).

3. You update your beliefs about \( \theta \) by combining information from the prior distribution and the data through the calculation of the **posterior distribution** \( \pi(\theta|\mathbf{x}) \).

The paradigm can be thought as a transformation from the before to the after:

\[
\pi(\theta) \rightarrow \pi(\theta|\mathbf{x})
\]
The Bayesian Method

Given data \( \mathbf{x} = \{x_1, \ldots, x_n\} \), Bayesian inference is carried out in the following way:

1. You choose a **prior distribution** \( \pi(\theta) \) for \( \theta \). The distribution describes your beliefs about the parameter *before* you examine the data.

2. Given the observed data \( \mathbf{x} \), you select a model (density) \( f(\mathbf{x}|\theta) \) to describe the distribution of \( \mathbf{x} \) given \( \theta \).

   The **Likelihood function** is any function proportional to \( f(\mathbf{x}|\theta) \); that is, \( \mathcal{L}(\theta) \propto f(\mathbf{x}|\theta) \).

3. You update your beliefs about \( \theta \) by combining information from the prior distribution and the data through the calculation of the **posterior distribution** \( \pi(\theta|\mathbf{x}) \).

The paradigm can be thought as a transformation from the before to the after:

\[
\pi(\theta) \longrightarrow \pi(\theta|\mathbf{x})
\]
Bayes’ Theorem

The updating of beliefs is carried out by using Bayes’ theorem, which enables you to combine the prior and the model in the following way:

\[
\pi(\theta|\mathbf{x}) = \frac{\pi(\theta, \mathbf{x})}{\pi(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\pi(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta) \, d\theta}
\]

The marginal distribution \(\pi(\mathbf{x})\) is an integral that is often ignored (as long as it is finite). Hence \(\pi(\theta|\mathbf{x})\) is often written as:

\[
\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) = \mathcal{L}(\theta)\pi(\theta)
\]

All inferences are based on the posterior distribution.
Bayesian Thinking in Real Life

You suspect you might have a fever and decide to take your temperature.

1. A possible prior density on your temperature $\theta$: likely normal (centered at 98.6) but possibly sick (centered at 101).

2. Suppose the thermometer says 101 degrees: $f(x|\theta) \sim N(\theta, \sigma^2)$ where $\sigma$ could be a very small number.

3. You get the posterior distribution. Yes, you are sick.
Inference

Some basic inferences are point estimation, interval estimation, and hypothesis testing.

- **Classical**
  - Estimator: MLE, MOME, least square estimator, and so on.

- **Bayesian**
  - All inference about $\theta$ is based on $\pi(\theta|x)$.
  - Estimator: mean, mode, median, any point from $\pi(\theta|x)$. For example, the posterior mean of $\theta$ is given by
    \[
    E(\theta|x) = \int_{\Theta} \theta \cdot \pi(\theta|x) d\theta
    \]
    The posterior mode of $\theta$ is the value of $\theta$ that maximizes $\pi(\theta|x)$. 
Interval Estimation

Classical:

- Estimator: confidence intervals
- Nice properties: coverage probability, minimize false coverage (uniformly most accurate), unbiasedness, and so on.
- A CI of $100(1 - \alpha)\%$ asserts that, in the long run, $100(1 - \alpha)\%$ of the realized confidence intervals cover the true parameter. You cannot say “The true parameter is in the CI with probability 95%.” The true parameter is either in or outside of a CI, not with any measurable probability.
- The interpretation reflects the uncertainty in the sampling procedure—the parameter is fixed, but the interval is random.
Interval Estimation

Bayesian:

- Estimator: credible sets
  any set $A$ such that $P(\theta \in A | x) = \int_A \pi(\theta | x) d\theta$

- Equal tail: $100(\alpha/2)$th and $100(1 - \alpha/2)$th percentiles.

- Highest posterior density (HPD):
  1. Posterior probability is $100(1 - \alpha)%$
  2. For $\theta_1 \in A$ and $\theta_2 \notin A$,
     $\pi(\theta_1 | x) \geq \pi(\theta_2 | x)$. The smallest region can be disjoint.

- Interpretation: “There is a 95% chance that the parameter is in this interval.” The parameter is random, not fixed.
Interval Estimation

Bayesian:

- Estimator: credible sets
  any set $A$ such that $P(\theta \in A|x) = \int_A \pi(\theta|x) \, d\theta$

- Equal tail: $100(\alpha/2)$th and $100(1 - \alpha/2)$th percentiles.

- Highest posterior density (HPD):
  1. Posterior probability is $100(1 - \alpha)$%
  2. For $\theta_1 \in A$ and $\theta_2 \notin A$,
     $\pi(\theta_1|x) \geq \pi(\theta_2|x)$. The smallest region can be disjoint.

- Interpretation: “There is a 95% chance that the parameter is in this interval.” The parameter is random, not fixed.
Hypothesis Testing

Classical:
- Methods of finding tests: likelihood ratio test, and so on.
- Method of evaluation: power, control Type I error, most powerful, and so on.

Bayesian:
- Test: \( P_{\Theta|X}(H_0|x)/P_{\Theta|X}(H_1|x) \) (posterior odds). Accept \( H_0 \) if \( P_{\Theta|X}(H_0|x)/P_{\Theta|X}(H_1|x) > c \).
- Posterior odds can be quite sensitive to the prior distributions. To correct for prior influence: \( \frac{P_{\Theta|X}(H_0|x)/P_{\Theta|X}(H_1|x)}{\pi_{\Theta}(H_0)/\pi_{\Theta}(H_1)} \). This is called the Bayes factor.
Outline

1. Introduction to Bayesian statistics
   - Background and concepts in Bayesian methods
   - Prior distributions
   - Advanced computational methods
   - Convergence diagnostics and summary statistics
   - Discussion and summary
Prior Distributions

The prior distribution represents your belief *before* seeing the data.

- Bayesian probability measures the degree of belief that you have in a random event. By this definition, probability is highly subjective. It follows that all priors are *subjective* priors.

- Not everyone agrees with the preceding. Some people would like to obtain results that are objectively valid, such as, “Let the data speak for itself.”. This approach advocates noninformative (flat/improper/Jeffreys) priors.

- Subjective approach advocates informative priors, which can be extraordinarily useful, if used correctly.
Noninformative Priors

- A prior is *noninformative* if it is *flat* relative to the likelihood function. Thus, a prior \( \pi(\theta) \) is noninformative if it has minimal impact on the posterior of \( \theta \).

- Many people like noninformative priors because they appear to be more objective. However, it is unrealistic to think that noninformative priors represent total ignorance about the parameter of interest.

- A frequent noninformative prior is \( \pi(\theta) \propto 1 \), which assigns equal likelihood to all possible values of the parameter.
A Binomial Example

- Suppose that you observe 14 heads in 17 tosses. The likelihood is:

  \[ \mathcal{L}(p) \propto p^x (1 - p)^{n-x} \]

  with \( x = 14 \) and \( n = 17 \).

- A flat prior on \( p \) is:

  \[ \pi(p) = 1 \]

- The posterior distribution is:

  \[ \pi(p|x) \propto p^{14} (1 - p)^3 \]

  which is a beta(15, 4).
Binomial with a Flat Prior

In the binomial example, the posterior distribution is identical to the likelihood function if a flat prior is used.
Flat Prior

If \( \pi(\theta|x) \propto \mathcal{L}(\theta) \) with \( \pi(\theta) \propto 1 \), then why not use the flat prior all the time? The short answer is:

- Using a flat prior does not always guarantee a proper (integrable) posterior distribution; that is, \( \int \pi(\theta|x)d\theta < \infty \).

The reason is that the likelihood function is only proper w.r.t. the random variable \( X \). But a posterior has to be integrable w.r.t. \( \theta \), a condition not required by the likelihood function.

In cases where the likelihood function and the posterior distribution are identical, very different approaches are used to carry out inferences.
Two Ways of Making Inferences

Classical inference typically uses asymptotic results; Bayesian inference is based on exploring the entire distribution.
Jeffreys’ Prior

Jeffreys’ prior is defined as

\[ \pi(\theta) \propto |I(\theta)|^{1/2} \]

where | · | denotes the determinant and \( I(\theta) \) is the expected Fisher information matrix based on the likelihood function \( p(x|\theta) \):

\[ I(\theta) = -E \left[ \frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} \right] \]
Jeffreys’ Prior, Binomial Example

The likelihood and log likelihood functions are:

\[ \mathcal{L}(p) \propto p^x (1 - p)^{n-x} \]

and

\[ \ell(p) = x \log(p) + (n - x) \log(1 - p) \]

The two derivatives are:

\[ \frac{\partial \ell}{\partial p} = \frac{x}{p} - \frac{n - x}{1 - p} \]

and

\[ \frac{\partial^2 \ell}{\partial p^2} = -\frac{x}{p^2} - \frac{n - x}{(1 - p)^2} \]
Jeffreys’ Prior, Binomial Example

The negative expected Fisher information, using $E(X) = np$, is

$$I(p) = \frac{-n}{p(1-p)}$$

Jeffreys’ prior is

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

The posterior distribution is

$$\mathcal{L}(p)\pi(p) \propto p^{x - \frac{1}{2}}(1-p)^{n-x - \frac{1}{2}}$$

$$\sim \text{Beta}(15.5, 4.5)$$
Jeffreys’ Prior

- Jeffreys
- Likelihood
- Posterior
Informative Priors

An informative prior is a prior that is not dominated by the likelihood and that has an impact on the posterior distribution.

- The proper use of prior distributions illustrates the power of the Bayesian method: information gathered from a previous study, past experience, or expert opinion can be combined with current information in a natural way.

- They are reasonable priors to use if one has real prior information from a previous similar study.

- Informative priors must be specified with care in actual practice. Otherwise, you can get misleading results.
Priors Can Be Too Informative!

![Graph showing prior, likelihood, and posterior distributions for temperature. The prior is too informative, causing the posterior distribution to be skewed towards the prior distribution.]
Outline

1 Introduction to Bayesian statistics
   - Background and concepts in Bayesian methods
   - Prior distributions
   - Advanced computational methods
   - Convergence diagnostics and summary statistics
   - Discussion and summary
Advanced Computational Methods

For many models, including GLMs, nonlinear models, random-effects models, survival models, and so on, the posterior distribution does not have a closed form.

In these cases, exact inference is not possible. You need to resort to approximation or sampling-based methods:

1. Asymptotic methods, like Laplace approximation
2. Direct simulation, such as inverse CDF
3. Importance sampling and rejection sampling
4. Stochastic simulation methods such as Markov chain Monte Carlo (MCMC)
Markov Chain Monte Carlo

- **Markov Chain**: a stochastic process that generates conditional independent samples according to some target distribution.
- **Monte Carlo**: a numerical integration technique that finds an expectation:
  \[
  E(f(\theta)) = \int f(\theta)p(\theta)d\theta \approx \frac{1}{n} \sum_{i=1}^{n} f(\theta_i)
  \]
  with \(\theta_1, \theta_2, \cdots, \theta_n\) being samples from \(p(\theta)\).
- **MCMC** is a method that generates a sequence of dependent samples from the target distribution and computes quantities by using Monte Carlo based on these samples.
- The most well-known algorithms are the Metropolis and Gibbs samplers.
Gibbs Sampler

**Gibbs sampler** is an algorithm that sequentially generates samples from a joint distribution of two or more random variables. The sampler is often used when:

- The joint distribution, $\pi(\theta|x)$, is not known explicitly
- The full conditional distribution of each parameter—for example, $\pi(\theta_i|\theta_j, i \neq j, x)$—is known
Gibbs Sampler

1. Set $t = 0$, and choose an arbitrary initial value of $\theta^{(0)} = \{\theta_1^{(0)}, \ldots, \theta_k^{(0)}\}$.

2. Generate each component of $\theta$ as follows:
   - Draw $\theta_1^{(t+1)}$ from $\pi(\theta_1 | \theta_2^{(t)}, \ldots, \theta_k^{(t)}, x)$
   - Draw $\theta_2^{(t+1)}$ from $\pi(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_k^{(t)}, x)$
   - ...  
   - Draw $\theta_k^{(t+1)}$ from $\pi(\theta_k | \theta_1^{(t+1)}, \ldots, \theta_{k-1}^{(t+1)}, x)$

3. Set $t = t + 1$. If $t < T$, the number of desired samples, return to step 2. Otherwise, stop.
Gibbs Sampler

\[ \pi(\alpha, \beta | x ) \]

\begin{figure}
\centering
\includegraphics[width=\textwidth]{gibbs_sampler_graph.png}
\end{figure}
Gibbs Sampler

\[ \pi(\alpha, \beta | x) \]

\[ \alpha^{(0)} \]

\[ \beta^{(0)} \]

\[ \pi(\beta | \alpha^{(0)}, x) \]
Gibbs Sampler

\[ \pi(\alpha, \beta | x) \]

\[ \pi(\alpha | \beta^{(0)}, x) \]

\[ \beta^{(0)} \]

\[ \alpha^{(1)} \quad \alpha^{(0)} \]

\[ \text{beta} \]

\[ \text{alpha} \]
Gibbs Sampler

\[ \pi(\alpha, \beta | x ) \]
Gibbs Sampler

\[ \pi(\alpha, \beta \mid x) \]

\(\alpha\) vs. \(\beta\) with a contour plot of the posterior distribution.
Gibbs Sampler

This works remarkably well in practice because:

- \( (\theta_1^{(t)}, \cdots, \theta_m^{(t)}) \xrightarrow{d} [\theta_1, \cdots, \theta_m] \) as \( t \to \infty \)

Under mild conditions (Besag 1974), the one-dimensional conditional distributions uniquely determine the full joint distribution \( \pi(\theta|x) \) and hence all marginal posterior distributions \( \pi(\theta_i|x) \).

- You can always find the (proportional) conditional distribution for any parameter:
  \[
  \pi(\theta_i|\theta_j, i \neq j, x) \propto p(x|\theta)\pi(\theta)
  \]

- Rejection sampling is a popular method to generate samples from low-dimensional (one-dimensional) distributions.
Outline

1 Introduction to Bayesian statistics
   - Background and concepts in Bayesian methods
   - Prior distributions
   - Advanced computational methods
   - Convergence diagnostics and summary statistics
   - Discussion and summary
Markov Chain Convergence

The importance of convergence diagnostics is obvious. An unconverged Markov chain does not explore the parameter space efficiently and the samples cannot approximate the target distribution well. Inference should not be based upon unconverged Markov chain, or very misleading results could be obtained.

It is important to remember:

- Convergence should be checked for ALL parameters, and not just those of interest.
- There are no definitive tests of convergence. Diagnostics are often not sufficient for convergence.
Convergence Terminology

- **Convergence**: initial drift in the samples towards a stationary (target) distribution
- **Burn-in**: samples at start of the chain that are discarded to minimize their impact on the posterior inference
- **Slow mixing**: tendency for high autocorrelation in the samples. A slow-mixing chain does not traverse the parameter space efficiently.
- **Thinning**: the practice of collecting every $k$th iteration to reduce autocorrelation. Thinning a Markov chain can be wasteful because you are throwing away a $\frac{k-1}{k}$ fraction of all the posterior samples generated.
- **Trace plot**: plot of sampled values of a parameter versus iteration number.
# Some Convergence Diagnostics Tests SAS Provides

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gelman-Rubin</td>
<td>Uses parallel chains with dispersed initial values to test whether they all converge to the same target distribution. Failure could indicate the presence of a multimode posterior distribution or the need to run a longer chain.</td>
<td>One-sided test based on a variance ratio test statistic. Large $\hat{R}_c$ values indicate rejection.</td>
</tr>
<tr>
<td>Geweke</td>
<td>Tests whether the mean estimates have converged by comparing means from the early and latter part of the Markov chain.</td>
<td>Two-sided test based on a $z$-score statistic. Large absolute $z$ values indicate rejection.</td>
</tr>
</tbody>
</table>
## Convergence Diagnostics Tests (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heidelberger-Welch stationarity test</td>
<td>Tests whether the Markov chain is a covariance (weakly) stationary process.</td>
<td>One-sided test based on a Cramer-von Mises statistic. Small $p$-values indicate rejection.</td>
</tr>
<tr>
<td>Heidelberger-Welch halfwidth test</td>
<td>Reports whether the sample size is adequate to meet the required accuracy for the mean estimate.</td>
<td>If relative half-width statistic is greater than a predetermined accuracy measure, this indicates rejection.</td>
</tr>
<tr>
<td>Raftery-Lewis</td>
<td>Evaluates the accuracy of the estimated (desired) percentiles by reporting the number of samples needed to reach the desired accuracy of the percentiles.</td>
<td>If the total samples needed are less than the Markov chain sample, this indicates rejection.</td>
</tr>
</tbody>
</table>
Graphical Display

Diagnostics for alpha

Iteration

Autocorrelation

Posterior Density

Lag

alpha
Graphical Display (Burn-In Needed?)
Graphical Display (Thinning Needed?)
Graphical Display (No Convergence)
More on Convergence Diagnosis

There are no definitive tests of convergence.

- *With experience*, visual inspection of trace plots is often the most useful approach.
- Geweke and Heidelberger-Welch sometimes reject 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.
How to Summarize Results from the MCMC Samples?

SAS Bayesian procedures report:

- Point estimates: mean, standard deviation, percentiles
- Interval estimates: equal-tail intervals and HPD intervals
- Posterior covariance and correlation matrices
- Effective sample sizes, Monte Carlo standard errors, fit statistics
- Kernel density estimation of the marginal posterior
- In some cases, summary statistics on functions of parameters, such as the hazard ratios.
Effective Sample Size

ESS is commonly used as a measure of how well a Markov chain is mixing.

\[
ESS = \frac{n}{1 + 2 \sum_{k=1}^{n-1} \rho_k(\theta)}
\]

where \( n \) is the total sample size and \( \rho_k(\theta) \) is the autocorrelation of lag \( k \) for \( \theta \).

The closer ESS is to \( n \), the better mixing is in the Markov chain.
Deviance Information Criterion

- DIC is a Bayesian alternative to AIC and BIC, a model assessment and selection tool.
- The criterion can be applied to non-nested models and models that have non-iid data.
- A smaller DIC indicates a better fit to the data.
Deviance Information Criterion (DIC)

\[ \text{DIC} = \bar{D}(\theta) + p_D = D(\bar{\theta}) + 2p_D \]

where

- \( D(\theta) = 2 \left( \log(f(y)) - \log(p(y|\theta)) \right) \) is the deviance where
  - \( p(y|\theta) \) is the likelihood function
  - \( f(y) \) is a constant term that is not calculated

- \( \bar{D}(\theta) \) is posterior mean of the deviance, approximated by \( \frac{1}{n} \sum_{t=1}^{n} D(\theta^t) \). The expected deviation measures how well the model fits the data.

- \( D(\bar{\theta}) \) is the deviance evaluated at \( \bar{\theta} \), equal to \( -2 \log(p(y|\bar{\theta})) \). It is the deviance evaluated at your “best” posterior estimate.

- \( p_D \) is the effective number of parameters.
Strengths of Bayesian Methods

- Provide a natural and principled way of combining prior information with data, within a solid decision-theoretical framework. All inferences logically follow from Bayes’ theorem and are based on the posterior distribution.
- Provide inferences that are conditional on the data and are exact, without reliance on either asymptotic approximation or the “plug-in” principle. Small sample inference proceeds in the same manner as if you had a large sample.
- Obey the likelihood principle.
Strengths of Bayesian Methods

- Provide interpretable answers, such as “the true parameter $\theta$ has a probability of 0.95 of falling in a 95% credible interval.” Bayesian analysis can answer specific scientific questions directly.

- Provide a convenient setting for a wide range of models, such as hierarchical models and missing data problems. MCMC, along with other numerical methods, makes computations tractable for virtually all parametric models.
Weaknesses of Bayesian Methods

- Do not tell you how to select a prior. Bayesian inferences require skills to translate subjective prior beliefs into a mathematically formulated prior. If you do not proceed with caution, you can generate misleading results.

- Can produce posterior distributions that are heavily influenced by the priors.

- Often come with a high computational cost, especially in models with a large number of parameters. In addition, simulations provide slightly different answers unless the same random seed is used.

- Do not produce credible intervals that are guaranteed to have the right coverage property, as the classical confidence intervals do.
A Short List of Bayesian Text Books


The “Introduction to Bayesian Analysis Procedures” section (in the *SAS/STAT® User’s Guide*) contains a “Bayesian Reading List” with comprehensive references.
Outline

1 Introduction to Bayesian statistics

2 The GENMOD, LIFEREG, and PHREG procedures
   - Overview
   - Prior distributions
   - The BAYES statement
   - GENMOD: linear regression
   - GENMOD: binomial model
   - PHREG: Cox model
   - PHREG: piecewise exponential model
   - LIFEREG: Weibull and exponential models (optional)
   - Summary
The GENMOD, LIFEREG, and PHREG Procedures

These three procedures provide:

- The BAYES statement (which carries out Bayesian analysis) as production in SAS 9.2
- The BAYES statement as experimental in SAS 9.1.3. (available as a Windows download)
- A set of frequently used prior distributions, posterior summary statistics, and convergence diagnostics
- Adaptive rejection (Gilks and Wild 1992; Gilks, Best, and Tan 1995) as a sampling method
GENMOD Procedure

PROC GENMOD provides Bayesian analysis for:

- **Distributions**: binomial, gamma, inverse-Gaussian, negative binomial, normal, and Poisson
- **Links**: identity, log, logit, probit, complementary log-log, and power

The procedure currently does not provide Bayesian analysis for:

- **Distributions**: multinomial
- **Links**: CCLL, CLogit, Cprobit

Model parameters are the regression coefficients and dispersion (or the precision or scale) parameter, if the model has one.
LIFEREG Procedure

PROC LIFEREG provides analyses for parametric lifetime models for:

**Distributions**: exponential, 3-parameter gamma, log-logistic, log-normal, logistic, normal, and Weibull

The procedure currently does not provide Bayesian analysis for the binomial distribution.

Model parameters are the regression coefficients and dispersion (or the precision or scale) parameter, if the model has one.
PHREG Procedure

PROC PHREG provides analysis for Bayesian semiparametric survival models:

- Cox regression models: use the partial likelihood as the likelihood (Sinha, Ibrahim, and Chen, 2003), time-independent and time-dependent, all TIES= methods
- Piecewise exponential models

The Bayesian functionality in PROC PHREG currently does not fit models with certain data constraints—for example, data that include recurrent events.

Model parameters are the regression coefficients and hazards (piecewise exponential models).
Outline

2 The GENMOD, LIFEREG, and PHREG procedures

- Overview
- Prior distributions
  - The BAYES statement
  - GENMOD: linear regression
  - GENMOD: binomial model
  - PHREG: Cox model
  - PHREG: piecewise exponential model
  - LIFEREG: Weibull and exponential models (optional)
- Summary
Prior Distributions in SAS Procedures

- **Uniform (or flat)** prior is defined as:
  \[ \pi(\theta) \propto 1 \]
  This prior is not integrable, but it does not lead to improper posterior in any of the procedures.

- **Improper** prior is defined as:
  \[ \pi(\theta) \propto \frac{1}{\theta} \]
  This prior is often used as a noninformative prior on the scale parameter, and it is uniform on the log-scale.

- **Proper** prior distributions include gamma, inverse-gamma, AR(1)-gamma, normal, multivariate normal densities.

- **Jeffreys’** prior is provided in PROC GENMOD.
GENMOD: Parameters and Priors

Priors supported:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression coefficients</td>
<td>Jeffreys’, normal, and uniform</td>
</tr>
<tr>
<td>dispersion</td>
<td>gamma, inverse-gamma, and improper</td>
</tr>
<tr>
<td>scale, precision</td>
<td>gamma, improper</td>
</tr>
</tbody>
</table>

The dispersion and regression parameters are assumed to be independent, except in the linear regression case.
LIFEREG: Parameters and Priors

Priors supported:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression coefficients</td>
<td>normal and uniform</td>
</tr>
<tr>
<td>scale</td>
<td>gamma and improper</td>
</tr>
<tr>
<td>3-parameter gamma shape</td>
<td>gamma and improper</td>
</tr>
</tbody>
</table>

The dispersion and regression parameters are assumed to be independent.
PHREG: Parameters and Priors

- **Piecewise exponential:**
  - Regression coefficients ($\beta$): normal and uniform priors
  - Log hazards ($\alpha = \log(\lambda)$): uniform and normal priors
  - Regression and log hazards: multivariate normal (do not need to be independent)
  - Hazards ($\lambda$): improper, uniform, independent gamma, and AR(1) priors

- **Cox:** uniform and normal priors on the regression coefficients
Outline

The GENMOD, LIFEREG, and PHREG procedures

- Overview
- Prior distributions
- The BAYES statement
- GENMOD: linear regression
- GENMOD: binomial model
- PHREG: Cox model
- PHREG: piecewise exponential model
- LIFEREG: Weibull and exponential models (optional)
- Summary
Syntax for the BAYES Statement

The BAYES statement is used to request all Bayesian analysis in these procedures.

```
BAYES < options > ;
```

The following options appear in all BAYES statements:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INITIAL=</td>
<td>initial values of the chain</td>
</tr>
<tr>
<td>NBI=</td>
<td>number of burn-in iterations</td>
</tr>
<tr>
<td>NMC=</td>
<td>number of iterations after burn-in</td>
</tr>
<tr>
<td>OUTPOST=</td>
<td>output data set for posterior samples</td>
</tr>
<tr>
<td>SEED=</td>
<td>random number generator seed seed</td>
</tr>
<tr>
<td>THINNING=</td>
<td>thinning of the Markov chain</td>
</tr>
<tr>
<td>DIAGNOSTICS=</td>
<td>convergence diagnostics</td>
</tr>
<tr>
<td>PLOTS=</td>
<td>diagnostic plots</td>
</tr>
<tr>
<td>SUMMARY=</td>
<td>summary statistics</td>
</tr>
<tr>
<td>COEFFPRIOR=</td>
<td>prior for the regression coefficients</td>
</tr>
</tbody>
</table>
The BAYES Statement

The following options are specific to PROC GENMOD:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISPERSIONPRIOR=</td>
<td>prior for the dispersion parameter</td>
</tr>
<tr>
<td>PRECISIONPRIOR=</td>
<td>prior for the precision parameter</td>
</tr>
<tr>
<td>SCALEPRIOR=</td>
<td>prior for the scale parameter</td>
</tr>
</tbody>
</table>
The BAYES Statement

The following options are specific to PROC LIFEREG:

<table>
<thead>
<tr>
<th>EXPONENTIALSCALEPRIOR=</th>
<th>prior for the exponential scale parm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCALEPRIOR=</td>
<td>prior for the scale parm</td>
</tr>
<tr>
<td>WEIBULLSCALEPRIOR=</td>
<td>prior for the Weibull scale parm</td>
</tr>
<tr>
<td>WEIBULLSHAPEPRIOR=</td>
<td>prior for the Weibull shape parm</td>
</tr>
<tr>
<td>GAMMASHAPEPRIOR=</td>
<td>prior for the gamma shape parm</td>
</tr>
</tbody>
</table>
The BAYES Statement

The following option is specific to PROC PHREG:

```
PIECEWISE= details of the piecewise exponential model
```

You can specify the number of intervals (number of hazards) or interval partitions and the prior distribution for the hazard parameters.
Outline

2 The GENMOD, LIFEREG, and PHREG procedures
- Overview
- Prior distributions
- The BAYES statement
- GENMOD: linear regression
  - GENMOD: binomial model
  - PHREG: Cox model
  - PHREG: piecewise exponential model
  - LIFEREG: Weibull and exponential models (optional)
- Summary
Regression Example

Consider the model

\[ Y = \beta_0 + \beta_1 \log X_1 + \epsilon \]

where \( Y \) is the survival time, \( \log X_1 \) is log(blood-clotting score), and \( \epsilon \) is a \( N(0, \sigma^2) \) error term.

The default priors that PROC GENMOD uses are:

\[ \pi(\beta_0) \propto 1 \quad \pi(\beta_1) \propto 1 \]
\[ \pi(\sigma) \sim \text{gamma}(\text{shape} = 0.001, \text{iscale} = 0.001) \]
Regression Example

A subset of the data:

data surg;
  input x1 logy @@;
  y = 10**logy;
  logx1 = log(x1);
datalines;
  6.7  2.3010  5.1  2.0043  7.4  2.3096
  6.5  2.0043  7.8  2.7067  5.8  1.9031
  5.7  1.9031  3.7  2.1038  6.0  2.3054
  3.7  2.3075  6.3  2.5172  6.7  1.8129
  ...
  5.1  2.1987  3.9  2.4914  6.6  2.0934
  6.4  2.0969  6.4  2.2967  8.8  2.4955
;
Regression Example

The following statements fit a Bayesian simple linear regression model:

```plaintext
ods graphics on;
proc genmod data=surg;
   model y = logx1 / dist=normal;
   bayes seed=4 outpost=post diagnostics=all summary=all;
run;
ods graphics off;
```

The SEED= option specifies a random seed; the OUTPOST= option saves posterior samples to the POST data set; the DIAGNOSTICS= and SUMMARY= options request the calculation for all convergence diagnostics and summary statistics.
Regression Example

PROC GENMOD provides some classical inference, such as the MLE.

Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-94.9822</td>
<td>114.5279</td>
<td>-319.453</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>129.4884</td>
</tr>
<tr>
<td>logx1</td>
<td>1</td>
<td>170.1749</td>
<td>65.8373</td>
<td>41.1361</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>299.2137</td>
</tr>
<tr>
<td>Scale</td>
<td>1</td>
<td>135.7963</td>
<td>13.0670</td>
<td>112.4556</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>163.9815</td>
</tr>
</tbody>
</table>

Note: The scale parameter was estimated by maximum likelihood.
Regression Example

Bayesian model information:

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Burn-In Size</td>
</tr>
<tr>
<td>MC Sample Size</td>
</tr>
<tr>
<td>Thinning</td>
</tr>
<tr>
<td>Distribution</td>
</tr>
<tr>
<td>Link Function</td>
</tr>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Survival Time</td>
</tr>
</tbody>
</table>
Regression Example

Priors on the coefficient and scale parameters:

### Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Constant</td>
</tr>
<tr>
<td>logx1</td>
<td>Constant</td>
</tr>
</tbody>
</table>

### Independent Prior Distributions for Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>Gamma</td>
<td>Shape: 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inverse Scale: 0.001</td>
</tr>
</tbody>
</table>
Convergence Diagnostics
Convergence Diagnostics

Diagnostics for logx1

![Graph showing diagnostics for logx1 with plots for lag, autocorrelation, iteration, and posterior density.](image)
Convergence Diagnostics

Diagnostics for Scale

Autocorrelation

Iteration

Posterior Density

Scale
Mixing

The following are the autocorrelation and effective sample sizes. The mixing appears to be very good, which agrees with the trace plots.

Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lag 1</th>
<th>Lag 5</th>
<th>Lag 10</th>
<th>Lag 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0085</td>
<td>-0.0063</td>
<td>-0.0124</td>
<td>-0.0101</td>
</tr>
<tr>
<td>logx1</td>
<td>0.0045</td>
<td>-0.0071</td>
<td>-0.0127</td>
<td>-0.0102</td>
</tr>
<tr>
<td>Scale</td>
<td>0.0371</td>
<td>0.0059</td>
<td>-0.0110</td>
<td>0.0141</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ESS</th>
<th>Correlation Time</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10000.0</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>logx1</td>
<td>10000.0</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Scale</td>
<td>9393.6</td>
<td>1.0646</td>
<td>0.9394</td>
</tr>
</tbody>
</table>
Convergence

The remaining convergence diagnostics tests all look good:

Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>97.5% Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>logx1</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>Scale</td>
<td>1.0003</td>
<td>1.0012</td>
</tr>
</tbody>
</table>

Raftery-Lewis Diagnostics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number of Samples</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Burn-in</td>
<td>Total</td>
<td>Minimum</td>
<td>Dependence Factor</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>3803</td>
<td>3746</td>
<td>1.0152</td>
</tr>
<tr>
<td>logx1</td>
<td>2</td>
<td>3932</td>
<td>3746</td>
<td>1.0497</td>
</tr>
<tr>
<td>Scale</td>
<td>2</td>
<td>3772</td>
<td>3746</td>
<td>1.0069</td>
</tr>
</tbody>
</table>
### Bayesian Analysis

| Parameter | z  | Pr > |z| |
|-----------|----|------|---|
| Intercept | -1.1206 | 0.2625 |
| logx1     | 1.1199  | 0.2627 |
| Scale     | 0.7145  | 0.4749 |

### Heidelberger-Welch Diagnostics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cramer-von-Mises Stat</th>
<th>p</th>
<th>Test Outcome</th>
<th>Iterations Discarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0788</td>
<td>0.6986</td>
<td>Passed</td>
<td>0</td>
</tr>
<tr>
<td>logx1</td>
<td>0.0544</td>
<td>0.8491</td>
<td>Passed</td>
<td>0</td>
</tr>
<tr>
<td>Scale</td>
<td>0.2176</td>
<td>0.2358</td>
<td>Passed</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Half-width</th>
<th>Mean</th>
<th>Relative Half-width</th>
<th>Test Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.9264</td>
<td>-95.6976</td>
<td>-0.0201</td>
<td>Passed</td>
</tr>
<tr>
<td>logx1</td>
<td>1.1139</td>
<td>170.7</td>
<td>0.00653</td>
<td>Passed</td>
</tr>
<tr>
<td>Scale</td>
<td>0.3369</td>
<td>140.2</td>
<td>0.00240</td>
<td>Passed</td>
</tr>
</tbody>
</table>
Summarize Convergence Diagnostics

- **Autocorrelation**: shows low dependency among Markov chain samples
- **ESS**: values close to the sample size indicate good mixing
- **Gelman-Rubin**: values close to 1 suggest convergence from different starting values
- **Geweke**: indicates mean estimates are stabilized
- **Raftery-Lewis**: shows sufficient samples to estimate 0.025 percentile within $+/- 0.005$ accuracy
- **Heidelberger-Welch**: suggests the chain has reached stationarity and there are enough samples to estimate the mean accurately
### Posterior Inference

Posterior summary and interval estimates:

#### Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>Intercept</td>
<td>10000</td>
<td>-95.6976</td>
<td>118.5</td>
<td>-173.8</td>
</tr>
<tr>
<td>logx1</td>
<td>10000</td>
<td>170.7</td>
<td>68.0965</td>
<td>124.8</td>
</tr>
<tr>
<td>Scale</td>
<td>10000</td>
<td>140.2</td>
<td>13.8824</td>
<td>130.4</td>
</tr>
</tbody>
</table>

#### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>-329.8</td>
<td>-333.2</td>
</tr>
<tr>
<td>logx1</td>
<td>0.050</td>
<td>39.7380</td>
<td>306.5</td>
</tr>
<tr>
<td>Scale</td>
<td>0.050</td>
<td>116.5</td>
<td>168.6</td>
</tr>
</tbody>
</table>
The GENMOD, LIFEREG, and PHREG procedures

**Fit Statistics**

PROC GENMOD also calculates various fit statistics (both classical and Bayesian). DIC is recommended for Bayesian analysis.

### Bayesian Analysis

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC (smaller is better)</td>
<td>689.650</td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>690.130</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>695.617</td>
</tr>
<tr>
<td>DIC (smaller is better)</td>
<td>689.689</td>
</tr>
<tr>
<td>pD (effective number of parameters)</td>
<td>2.966</td>
</tr>
</tbody>
</table>
Posterior Probabilities

Suppose that you are interested in knowing whether $\log X_1$ has a positive effect on survival time. Quantifying that measurement, you can calculate the probability $\beta_1 > 0$, which can be estimated directly from the posterior samples:

$$Pr(\beta_1 > 0 | Y, \log X_1) = \frac{1}{N} \sum_{t=1}^{N} I(\beta_1^t > 0)$$

where $I(\beta_1^t > 0) = 1$ if $\beta_1^t > 0$ and 0 otherwise. $N = 10,000$ is the sample size in this example.
Posterior Probabilities

The following SAS statements calculate the posterior probability:

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

ods select summary;
proc means data = Prob(keep=Indicator) n mean;
run;
```

The probability is roughly 0.9926, which strongly suggests that the slope coefficient is greater than 0.
Outline

2 The GENMOD, LIFEREG, and PHREG procedures
   • Overview
   • Prior distributions
   • The BAYES statement
   • GENMOD: linear regression
   • GENMOD: binomial model
   • PHREG: Cox model
   • PHREG: piecewise exponential model
   • LIFEREG: Weibull and exponential models (optional)
   • Summary
Binomial model

Researchers are interested in evaluating the performance of a medical procedure in a multicenter study. The following statements create a SAS data set for the treatment arm of the trials:

```sas
data trials;
  input event n;
datalines;
  2   86
  2   69
  1   71
  1  113
  1  103
;
```

**event:** number of deaths

**n:** number of patients assigned to the treatment procedure
Binomial Example

Consider a simple binomial model

\[
event_i \sim \text{binomial}(n_i, p)
\]
\[
p \sim \text{beta}(a, b)
\]

where \( p \) is the parameter of interest and \( a \) and \( b \) are hyper-parameters. Consider the following choices for \( a \) and \( b \):

- Jeffreys’: \( \text{beta}(-0.5, -0.5) \).
- uniform: \( \text{beta}(1, 1) \)

You know the posterior distribution of \( p \) in closed-form.
Noninformative Priors in the Binomial Example

The uniform and Jeffreys’ prior distributions, with corresponding posterior distributions, in the binomial example.
Binomial Example: Jeffreys’ Prior

To fit the model in PROC GENMOD, use the following transformation where the parameter of interest is the intercept:

\[ p = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \]

The prior on \( p \) needs to be transformed to the \( \beta_0 \) parameterization. Jeffreys’ prior is invariant to transformation. You can use the Jeffreys’ prior on the regression coefficient in PROC GENMOD:

```
proc genmod data=trials;
  model event/n= / dist=b link=logit;
  bayes seed=7 outpost=bout1 cprior=jeffreys;
run;
```
Binomial Example: Jeffreys’ Prior

Posterior summary statistics of $\beta_0$:

**Bayesian Analysis**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10000</td>
<td>-4.1282</td>
<td>0.3793</td>
<td>-4.3692</td>
<td>-4.1066</td>
<td>-3.8623</td>
</tr>
</tbody>
</table>

**Posterior Intervals**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>-4.9215</td>
<td>-3.4548</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>-4.9215</td>
<td>-3.4548</td>
</tr>
</tbody>
</table>
Convergence Diagnostics

Diagnostics for Intercept

Autocorrelation vs. Lag

Posterior Density vs. Intercept
Binomial Example: Jeffreys’ Prior

Transforming the posterior samples of $\beta_0$ to $p$ is relatively straightforward. Use the logistic function in the DATA step:

```plaintext
data b1 (keep=iteration jefp);
  set bout1;
  jefp = logistic(b);
  output;
run;
```
Binomial Example: Uniform Prior

Uniform prior on $p$ corresponds to a logistic prior on $\beta_0$:

$$p \sim \text{beta}(1, 1) \Leftrightarrow \pi(\beta_0) = \frac{\exp(-\beta_0)}{(1 + \exp(-\beta_0))^2}$$

PROC GENMOD does not allow a logistic prior on the regression coefficient, but a normal prior gives a good approximation:

```plaintext
proc genmod data=trials;
  model event/n= / dist=b
  link=logit;
  bayes seed=7 outpost=bout2
cprior=normal(var=3.289868);
run;
```
You start with

\[ p = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \frac{1}{1 + \exp(-\beta_0)} \]

\[ \frac{\partial p}{\partial \beta_0} = -\frac{\exp(-\beta_0)}{(1 + \exp(-\beta_0))^2} \]

Do the transformation of variables, with the Jacobian:

\[ f(p) = 1 \cdot I_{\{0 \leq p \leq 1\}} \]

\[ \Rightarrow f(\beta_0) = \left| \frac{\partial p}{\partial \beta_0} \right| \cdot I_{\{0 \leq \frac{1}{1 + \exp(-\beta_0)} \leq 1\}} = \frac{\exp(-\beta_0)}{(1 + \exp(-\beta_0))^2} \cdot I_{\{-\infty \leq \beta_0 \leq \infty\}} \]

The pdf for the logistic distribution with location \( a \) and scale \( b \) is

\[ \exp\left(-\frac{\beta_0 - a}{b}\right) \bigg/ b \left(1 + \exp\left(-\frac{\beta_0 - a}{b}\right)\right)^2 \]

with mean \( a \) and variance \( \frac{\pi^2 b^2}{3} \). You set \( a = 0 \) and \( b = 1 \) to get the standard logistic distribution.
Simulation Results

Estimated posterior distributions of $p$ using PROC GENMOD. The estimated densities closely resemble the true densities, with minor deviation in the normal approximation prior case.
Some Thoughts

It is difficult to be truly noninformative: uniform prior on $p$ and uniform prior on $\beta_0$ lead to different posterior distributions.
Some Thoughts

Jeffreys’ prior is

- *locally uniform*—a prior that does not change much over the region in which the likelihood is significant and does not assume large values outside that range. Hence it is somewhat noninformative.
- invariant with respect to one-to-one transformations.

The prior also

- violates the likelihood principle
- can be improper for many models
- can be difficult to construct

Jeffreys’ priors in PROC GENMOD does not lead to improper posteriors.
Other values of $a$ and $b$?

Empirical Bayes’ offers one solution: use the data to estimate $a$ and $b$. It involves optimization w.r.t. the marginal posterior distribution of the hyperparameters, given data (with $p$ integrated out). This distribution is beta-binomial:

$$
\pi(a, b|x, n) = \binom{n}{x} \frac{\Gamma(b + n - x) \Gamma(a + x) \Gamma(a + b)}{\Gamma(a + b + n) \Gamma(a) \Gamma(b)}
$$
Empirical Bayes Approach

Optimization can be carried out, on its log-scale, using PROC NLMIXED.

```
proc nlmixed data=trials;
   parms a 40 b 2000;
   lf = lgamma(b+n-event) + lgamma(a+event) + lgamma(a+b)
       - lgamma(a+b+n) - lgamma(a) - lgamma(b);
   model event ~ general(lf);
run;
```

It turns out that this objective function is sensitive to initial values, and has a very flat surface on a large support. Nevertheless, all converged estimates result in small $a$ and large $b$. In this run, we have $\hat{a} = 32.6178$ and $\hat{b} = 2000.18$. 
This Prior Might Be Too Informative

The Empirical Bayes prior, beta(32.6, 2000), dominates the likelihood function.
Other Options?

You can place another layer of prior distribution on the hyperparameters $a$ and $b$. PROC GENMOD does not have this capability in handling multilevel hierarchical models. You can do that with PROC MCMC.

Another possibility is, if a pilot study or historical data is available, construct a more informative prior, such as the *power prior*, in the analysis.
Outline

The GENMOD, LIFEREG, and PHREG procedures

- Overview
- Prior distributions
- The BAYES statement
- GENMOD: linear regression
- GENMOD: binomial model
- PHREG: Cox model
- PHREG: piecewise exponential model
- LIFEREG: Weibull and exponential models (optional)
- Summary
Cox Model

Consider the data for the Veterans Administration lung cancer trial presented in Appendix 1 of Kalbfleisch and Prentice (1980).

<table>
<thead>
<tr>
<th>Time</th>
<th>Death in days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapy</td>
<td>Type of therapy: standard or test</td>
</tr>
<tr>
<td>Cell</td>
<td>Type of tumor cell: adeno, large, small, or squamous</td>
</tr>
<tr>
<td>PTherapy</td>
<td>Prior therapy: yes or no</td>
</tr>
<tr>
<td>Age</td>
<td>Age in years</td>
</tr>
<tr>
<td>Duration</td>
<td>Months from diagnosis to randomization</td>
</tr>
<tr>
<td>KPS</td>
<td>Karnofsky performance scale</td>
</tr>
<tr>
<td>Status</td>
<td>Censoring indicator (1=censored time, 0=event time)</td>
</tr>
</tbody>
</table>
## Cox Model

A subset of the data:

<table>
<thead>
<tr>
<th>OBS</th>
<th>Therapy</th>
<th>Cell</th>
<th>Time</th>
<th>Kps</th>
<th>Duration</th>
<th>Age</th>
<th>Ptherapy</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>standard</td>
<td>squamous</td>
<td>72</td>
<td>60</td>
<td>7</td>
<td>69</td>
<td>no</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>standard</td>
<td>squamous</td>
<td>411</td>
<td>70</td>
<td>5</td>
<td>64</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>standard</td>
<td>squamous</td>
<td>228</td>
<td>60</td>
<td>3</td>
<td>38</td>
<td>no</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>standard</td>
<td>squamous</td>
<td>126</td>
<td>60</td>
<td>9</td>
<td>63</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>standard</td>
<td>squamous</td>
<td>118</td>
<td>70</td>
<td>11</td>
<td>65</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>standard</td>
<td>squamous</td>
<td>10</td>
<td>20</td>
<td>5</td>
<td>49</td>
<td>no</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>standard</td>
<td>squamous</td>
<td>82</td>
<td>40</td>
<td>10</td>
<td>69</td>
<td>yes</td>
<td>1</td>
</tr>
</tbody>
</table>

...
Cox Model

- Some parameters are the coefficients of the continuous variables (KPS, Duration, and Age).
- Other parameters are the coefficients of the design variables for the categorical explanatory variables (PTherapy, Cell, and Therapy).
- You can use the CLASS statement in PROC PHREG to specify the categorical variables and their reference levels, such as CLASS PTherapy(ref='no').
**Cox Model**

The model considered here is the Breslow partial likelihood:

\[
L(\beta) = \prod_{i=1}^{k} \frac{e^{\beta' \sum_{j \in D_i} Z_j(t_i)}}{\left[ \sum_{l \in R_i} e^{\beta' Z_l(t_i)} \right]^{d_i}}
\]

where

- \( t_1 < \cdots < t_k \) are distinct event times
- \( Z_j(t_i) \) is the vector explanatory variables for the \( j \)th individual at time \( t_i \)
- \( R_i \) is the risk set at \( t_i \), which includes all observations that have survival time greater than or equal to \( t_i \)
- \( d_i \) is the multiplicity of failures at \( t_i \). It is the size of the set \( D_i \) of individuals that fail at \( t_i \)
Cox Model

The following statements fit a Cox regression model with a uniform prior on the regression coefficients:

```proc phreg data=VALung;
    class PTherapy(ref='no') Cell(ref='large') Therapy(ref='standard');
    model Time*Status(0) = KPS Duration Age PTherapy Cell Therapy;
    bayes seed=1 outpost=cout coeffprior=uniform plots=density;
run;```
Cox Model

Posterior Density Plots

Kps

Duration

Age

Ptherapyes
Cox Model

Posterior Density Plots

Celladeno

Cellsmall

Cellsquamous

Therapytest
Cox Model: Posterior Mean Estimates

### Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kps</td>
<td>10000</td>
<td>-0.0327</td>
<td>0.00545</td>
<td>-0.0364</td>
<td>-0.0328</td>
<td>-0.0291</td>
</tr>
<tr>
<td>Duration</td>
<td>10000</td>
<td>-0.00170</td>
<td>0.00945</td>
<td>-0.00791</td>
<td>-0.00123</td>
<td>0.00489</td>
</tr>
<tr>
<td>Age</td>
<td>10000</td>
<td>-0.00852</td>
<td>0.00935</td>
<td>-0.0147</td>
<td>-0.00850</td>
<td>-0.00223</td>
</tr>
<tr>
<td>Ptherapyyes</td>
<td>10000</td>
<td>0.0754</td>
<td>0.2345</td>
<td>-0.0776</td>
<td>0.0766</td>
<td>0.2340</td>
</tr>
<tr>
<td>Celladeno</td>
<td>10000</td>
<td>0.7867</td>
<td>0.3080</td>
<td>0.5764</td>
<td>0.7815</td>
<td>0.9940</td>
</tr>
<tr>
<td>Cellsmall</td>
<td>10000</td>
<td>0.4632</td>
<td>0.2731</td>
<td>0.2775</td>
<td>0.4602</td>
<td>0.6435</td>
</tr>
<tr>
<td>Cellsquamous</td>
<td>10000</td>
<td>-0.4022</td>
<td>0.2843</td>
<td>-0.5935</td>
<td>-0.4024</td>
<td>-0.2124</td>
</tr>
<tr>
<td>Therapytest</td>
<td>10000</td>
<td>0.2897</td>
<td>0.2091</td>
<td>0.1500</td>
<td>0.2900</td>
<td>0.4294</td>
</tr>
</tbody>
</table>
Cox Model: Interval Estimates

**Bayesian Analysis**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kps</td>
<td>0.050</td>
<td>-0.0433, -0.0219</td>
<td>-0.0434, -0.0221</td>
</tr>
<tr>
<td>Duration</td>
<td>0.050</td>
<td>-0.0216, 0.0153</td>
<td>-0.0202, 0.0164</td>
</tr>
<tr>
<td>Age</td>
<td>0.050</td>
<td>-0.0271, 0.00980</td>
<td>-0.0270, 0.00983</td>
</tr>
<tr>
<td>Ptherapyyes</td>
<td>0.050</td>
<td>-0.3943, 0.5335</td>
<td>-0.3715, 0.5488</td>
</tr>
<tr>
<td>Celladeno</td>
<td>0.050</td>
<td>0.1905, 1.3969</td>
<td>0.1579, 1.3587</td>
</tr>
<tr>
<td>Cellsmall</td>
<td>0.050</td>
<td>-0.0617, 1.0039</td>
<td>-0.0530, 1.0118</td>
</tr>
<tr>
<td>Cellsquamous</td>
<td>0.050</td>
<td>-0.9651, 0.1519</td>
<td>-0.9550, 0.1582</td>
</tr>
<tr>
<td>Therapytest</td>
<td>0.050</td>
<td>-0.1191, 0.6955</td>
<td>-0.1144, 0.6987</td>
</tr>
</tbody>
</table>
Cox Model: Plotting Survival Curves

Suppose that you are interested in estimating the survival curves for two individuals who have similar characteristics, with one receiving the standard treatment while the other did not. The following is saved in the SAS data set `pred`:

<table>
<thead>
<tr>
<th>OBS</th>
<th>Ptherapy</th>
<th>kps</th>
<th>duration</th>
<th>age</th>
<th>cell</th>
<th>therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no</td>
<td>58</td>
<td>8.7</td>
<td>60</td>
<td>large</td>
<td>standard</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>58</td>
<td>8.7</td>
<td>60</td>
<td>large</td>
<td>test</td>
</tr>
</tbody>
</table>
Cox Model

You can use the following statements to estimate the survival curves and save the estimates to a SAS data set:

```sas
proc phreg data=VALung plots(cl=hpd overlay)=survival;
    baseline covariates=pred out=pout;
    class PTherapy(ref='no') Cell(ref='large')
        Therapy(ref='standard');
    model Time*Status(0) = KPS Duration Age PTherapy Cell Therapy;
    bayes seed=1 outpost=cout coeffprior=uniform
        plots=density;
run;
```
Cox Model: Posterior Survival Curves

Estimated survival curves for the two subjects and their corresponding 95% HPD intervals.
Hazard Ratios

Hazard ratio, a ratio of two hazard functions, is a random variable—functions of the parameters are random variables themselves. Therefore you can get the posterior distributions for any hazard ratios and make inference by using the distributions. The new HAZARDRATIO statement enables you to obtain customized hazard ratios.

HAZARDRATIO '<label>' variables < / options > ;

- For a continuous variable: the hazard ratio compares the hazards for a given change (by default, a increase of 1 unit) in the variable.
- For a CLASS variable, a hazard ratio compares the hazards of two levels of the variable.
- This is a new statement you can use for both classical and Bayesian analyses.
Hazard Ratios

The following SAS statements fit the same Cox regression model and request three kinds of hazard ratios.

```sas
proc phreg data=VALung;
  class PTherapy(ref='no') Cell(ref='large') Therapy(ref='standard');
  model Time*Status(0) = KPS Duration Age PTherapy Cell Therapy;
  bayes seed=1 outpost=vout plots=trace coeffprior=uniform;
  hazardratio 'HR 1' Therapy / at(PTherapy='yes' KPS=80 duration=12 age=65 cell='small');
  hazardratio 'HR 2' Age / unit=10 at(KPS=45);
  hazardratio 'HR 3' Cell;
run;
```
Hazard Ratios
The following results are the summary statistics of the posterior hazards between the standard therapy and the test therapy.

Bayesian Analysis

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapy standard vs test At Prior=yes Kps=80 Duration=12 Age=65 Cell=small</td>
<td>10000</td>
<td>0.7651</td>
<td>0.1617</td>
<td>0.6509</td>
<td>0.7483</td>
<td>0.8607</td>
</tr>
</tbody>
</table>

HR 1: Hazard Ratios for Therapy

<table>
<thead>
<tr>
<th>95% Equal-Tail Interval</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4988</td>
<td>0.4692</td>
</tr>
<tr>
<td>1.1265</td>
<td>1.0859</td>
</tr>
</tbody>
</table>
Hazard Ratios

The following table lists the change of hazards for an increase in Age of 10 years.

Bayesian Analysis

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95% Equal-Tail Interval</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Unit=10 At Kps=45</td>
<td>10000</td>
<td>0.9224</td>
<td>0.0865</td>
<td>0.8633</td>
<td>0.9185</td>
<td>0.9779</td>
<td>0.7629</td>
<td>1.1030</td>
</tr>
</tbody>
</table>
Hazard Ratios

The following table lists posterior hazards between different levels in the Cell variable:

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95% Equal-Tail Interval</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell adeno vs large</td>
<td>10000</td>
<td>2.3035</td>
<td>0.7355</td>
<td>1.7797</td>
<td>2.1848</td>
<td>2.7020</td>
<td>1.2099 4.0428</td>
<td>1.0661 3.7509</td>
</tr>
<tr>
<td>Cell adeno vs small</td>
<td>10000</td>
<td>1.4374</td>
<td>0.4124</td>
<td>1.1479</td>
<td>1.3811</td>
<td>1.6622</td>
<td>0.7985 2.3857</td>
<td>0.7047 2.2312</td>
</tr>
<tr>
<td>Cell adeno vs squamous</td>
<td>10000</td>
<td>3.4376</td>
<td>1.0682</td>
<td>2.6679</td>
<td>3.2903</td>
<td>4.0199</td>
<td>1.8150 5.9733</td>
<td>1.6274 5.6019</td>
</tr>
<tr>
<td>Cell large vs small</td>
<td>10000</td>
<td>0.6530</td>
<td>0.1798</td>
<td>0.5254</td>
<td>0.6311</td>
<td>0.7577</td>
<td>0.3664 1.0636</td>
<td>0.3357 1.0141</td>
</tr>
<tr>
<td>Cell large vs squamous</td>
<td>10000</td>
<td>1.5567</td>
<td>0.4514</td>
<td>1.2367</td>
<td>1.4954</td>
<td>1.8103</td>
<td>0.8591 2.6251</td>
<td>0.7776 2.4679</td>
</tr>
<tr>
<td>Cell small vs squamous</td>
<td>10000</td>
<td>2.4696</td>
<td>0.7046</td>
<td>1.9717</td>
<td>2.3742</td>
<td>2.8492</td>
<td>1.3872 4.1403</td>
<td>1.2958 3.9351</td>
</tr>
</tbody>
</table>
Outline

2 The GENMOD, LIFEREG, and PHREG procedures
   - Overview
   - Prior distributions
   - The BAYES statement
   - GENMOD: linear regression
   - GENMOD: binomial model
   - PHREG: Cox model
   - PHREG: piecewise exponential model
   - LIFEREG: Weibull and exponential models (optional)
   - Summary
Piecewise Exponential Model

Let \( \{(t_i, x_i, \delta_i), i = 1, 2, \ldots, n\} \) be the observed data. Let \( a_0 = 0 < a_1 < \ldots < a_{J-1} < a_J = \infty \) be a partition of the time axis. The hazard for subject \( i \) is

\[
h(t|x_i; \theta) = h_0(t) \exp(\beta'x_i)
\]

where

\[
h_0(t) = \lambda_j \quad a_{j-1} \leq t < a_j \quad (j = 1, \ldots, J)
\]

The hazard for subject \( i \) in the \( j \)th time interval is

\[
h(t) = \lambda_j \exp(\beta'x_i) \quad a_{j-1} < t < a_j
\]
Piecewise Exponential Model

From the hazard function, first define the baseline cumulative hazard function:

$$H_0(t) = \sum_{j=1}^{J} \lambda_j \Delta_j(t)$$

where

$$\Delta_j(t) = \begin{cases} 
0 & t < a_{j-1} \\
 t - a_{j-1} & a_{j-1} \leq t < a_j \\
 a_j - a_{j-1} & t \geq a_j 
\end{cases}$$
**Piecewise Exponential Model**

The log likelihood is:

\[
l(\lambda, \beta) = \sum_{i=1}^{n} \delta_i \left[ \sum_{j=1}^{J} I(a_{j-1} \leq t_i < a_j) \log \lambda_j + \beta' x_i \right] - \sum_{i=1}^{n} \left[ \sum_{j=1}^{J} \Delta_j(t_i)\lambda_j \right] \exp(\beta' x_i)
\]

where \( \delta_i \) is the event status:

\[
\delta_i = \begin{cases} 
0 & \text{if } t_i \text{ is a censored time} \\
1 & \text{if } t_i \text{ is an event time}
\end{cases}
\]

This model has two parameter vectors: \( \lambda \) and \( \beta \).
PROC PHREG supports the following priors for the piecewise exponential model:

- Regression coefficients ($\beta$): normal and uniform priors
- Hazards ($\lambda$): improper, uniform, independent gamma, and AR(1) priors
- Log hazards ($\alpha = \log(\lambda)$): uniform and normal priors
- Regression coefficients and log hazards: multivariate normal (do not need to be independent)
**Piecewise Exponential Model**

For the hazard parameter $\lambda$, you can specify the following priors:

- **Improper:**
  $$
  \pi(\lambda) \propto \prod_{j=1}^{J} \frac{1}{\lambda_j}
  $$

- **Uniform:**
  $$
  \pi(\lambda) \propto 1
  $$

- **Independent gamma:**
  $$
  \pi(\lambda) \propto \prod_{j=1}^{J} \left\{ \lambda_j^{a_j-1} \exp(-\lambda_j b_j) \right\}
  $$
Piecewise Exponential Model

The AR(1) gamma prior for $\lambda_1, \ldots, \lambda_J$ is given by:

\[
\begin{align*}
\lambda_1 & \sim G(a_1, b_1) \\
\lambda_2 | \lambda_1 & \sim G\left(a_2, \frac{b_2}{\lambda_1}\right) \\
\vdots & \quad \vdots \\
\lambda_J | \lambda_{J-1} & \sim G\left(a_J, \frac{b_J}{\lambda_{J-1}}\right)
\end{align*}
\]

The joint prior density is given by:

\[
p(\lambda_1, \ldots, \lambda_J) \propto \lambda_1^{a_1-1} \exp(-b_1 \lambda_1) \prod_{j=2}^{J} \left(\frac{b_j}{\lambda_{j-1}}\right)^{a_j} \lambda_j^{a_j-1} \exp\left(-\frac{b_j}{\lambda_{j-1}} \lambda_j\right)
\]
Piecewise Exponential Model

For the logarithm of hazard, $\alpha = \log(\lambda)$, you can specify:

- **Uniform:**
  \[
  \pi(\alpha) \propto 1
  \]

- **Multivariate normal:**
  \[
  \pi(\alpha) \propto \exp\left[-\frac{1}{2}(\alpha - \alpha_0)'\Phi_0^{-1}(\alpha - \alpha_0)\right]
  \]

- You can specify a joint multivariate normal prior for $(\alpha, \beta)$:
  \[
  \pi(\alpha, \beta) \propto \exp\left[-\frac{1}{2}[(\alpha - \alpha_0)', (\beta - \beta_0)']\Sigma_0^{-1}[(\alpha - \alpha_0)', (\beta - \beta_0)']\right]
  \]
Consider a randomized trial of 40 rats exposed to carcinogen:

- Drug X and Placebo are the treatment groups.
- Event of interest is death.
- Response is time until death.
- What are the effects of treatment and gender on survival?
Piecewise Exponential Model

A subset of the data:

```sas
proc format;
   value Rx 1='X' 0='Placebo';
data Exposed;
   input Days Status Trt Gender $ @@;
   format Trt Rx.;
datalines;
179  1  1  F  378  0  1  M
256  1  1  F  355  1  1  M
262  1  1  M  319  1  1  M
256  1  1  F  256  1  1  M
...  
268  0  0  M  209  1  0  F
;
```
Piecewise Exponential Model

An appropriate model is the piecewise exponential. In the model:

- Each time interval has a constant hazard
- There are a total of eight intervals (PROC PHREG default)
- Intervals are determined by placing roughly equal number of uncensored observations in each interval
- The log hazard is used. It is generally more computationally stable. There are 8 $\lambda_i$'s and two regression coefficients.
Piecewise Exponential Model

The following programming statements fit a Bayesian piecewise exponential model with noninformative priors on both $\beta$ and $\log(\lambda)$:

```plaintext
proc phreg data=Exposed;
  class Trt(ref='Placebo') Gender(ref='F');
  model Days*Status(0)=Trt Gender;
  bayes seed=1 outpost=eout piecewise=loghazard(n=8);
run;
```

The PIECEWISE= option requests the estimating of a piecewise exponential model with 8 intervals.
Piecewise Exponential Model

Suppose that you have some prior information w.r.t. both $\beta$ and $\log(\lambda)$ that can be approximated well with a multivariate normal distribution. You can construct the following data set:

data pinfo;
  input _TYPE_ $ alpha1-alpha8 trtX GenderM;
datalines;
  Mean  0 0 0 0 0 0 0 0 0 0
  cov  90.2 -9.8  1.3 -1.9  4.1  3.7 14.3 -10.7 -7.2 -4.2
  cov -9.8 102.4 15.3 -12.1 15.6  6.8 -23.7 -23.7  9.0 -8.8
  cov  1.3  15.3 102.8  13.0 22.1  5.7 21.4 -16.1 14.2  13.3
  cov -1.9 -12.1  13.0  90.2  4.6 -16.1 11.3  -8.6 -12.6 -1.2
  cov  4.1  15.6  22.1  4.6 107.9 18.2  2.4 -8.1  2.9 -16.4
  cov  3.7  6.8  5.7 -16.1 18.2 123.3 -2.7 -7.9  3.2 -3.4
  cov 14.3 -23.7  21.4  11.3  2.4 -2.7 114.2  2.3  6.7 11.6
  cov -10.7 -23.7 -16.1 -8.6 -8.1 -7.9  2.3 91.8 -7.6  0.0
  cov -7.2  9.0 14.2 -12.6  2.9  3.2  6.7 -7.6 100.0 -6.3
  cov -4.2 -8.8 13.3 -1.2 -16.4 -3.4 11.6  0.0 -6.3 124.7;

Pieznwise Exponential Model

The following programming statements fit a Bayesian piecewise exponential model with informative prior on both $\beta$ and $\log(\lambda)$:

```plaintext
proc phreg data=exposed;
  class Trt(ref='Placebo') Gender(ref='F');
  model Days*Status(0)=Trt Gender;
  bayes seed=1 outpost=eout
    piecewise=loghazard(n=8 prior=normal(input=pinfo))
    cprior=normal(input=pinfo);
run;
```
Piecewise Exponential Model (Noninformative Analysis)

Bayesian Analysis

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Censoring Variable</td>
</tr>
<tr>
<td>Censoring Value(s)</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Burn-In Size</td>
</tr>
<tr>
<td>MC Sample Size</td>
</tr>
<tr>
<td>Thinning</td>
</tr>
</tbody>
</table>

Summary of the Number of Event and Censored Values

<table>
<thead>
<tr>
<th>Total</th>
<th>Event</th>
<th>Censored</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>36</td>
<td>4</td>
<td>10.00</td>
</tr>
</tbody>
</table>
### Piecewise Exponential Model

The partition of the time intervals:

<table>
<thead>
<tr>
<th>Interval</th>
<th>N</th>
<th>Event</th>
<th>Log Hazard Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0, 193)</td>
<td>193</td>
<td>5</td>
<td>5 Alpha1</td>
</tr>
<tr>
<td>193, 221</td>
<td>221</td>
<td>5</td>
<td>5 Alpha2</td>
</tr>
<tr>
<td>221, 239.5</td>
<td>239.5</td>
<td>7</td>
<td>5 Alpha3</td>
</tr>
<tr>
<td>239.5, 255.5</td>
<td>255.5</td>
<td>5</td>
<td>5 Alpha4</td>
</tr>
<tr>
<td>255.5, 256.5</td>
<td>256.5</td>
<td>4</td>
<td>4 Alpha5</td>
</tr>
<tr>
<td>256.5, 278.5</td>
<td>278.5</td>
<td>5</td>
<td>4 Alpha6</td>
</tr>
<tr>
<td>278.5, 321</td>
<td>321</td>
<td>4</td>
<td>4 Alpha7</td>
</tr>
<tr>
<td>321, Infty</td>
<td>Infty</td>
<td>5</td>
<td>4 Alpha8</td>
</tr>
</tbody>
</table>
### Piecewise Exponential Model

Posterior summary statistics:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha1</td>
<td>10000</td>
<td>-6.4137</td>
<td>0.4750</td>
<td>-6.7077</td>
<td>-6.3770</td>
<td>-6.0852</td>
</tr>
<tr>
<td>Alpha2</td>
<td>10000</td>
<td>-4.0505</td>
<td>0.4870</td>
<td>-4.3592</td>
<td>-4.0207</td>
<td>-3.7058</td>
</tr>
<tr>
<td>Alpha3</td>
<td>10000</td>
<td>-2.9297</td>
<td>0.5146</td>
<td>-3.2468</td>
<td>-2.8954</td>
<td>-2.5737</td>
</tr>
<tr>
<td>Alpha4</td>
<td>10000</td>
<td>-1.9146</td>
<td>0.6212</td>
<td>-2.3256</td>
<td>-1.8936</td>
<td>-1.4839</td>
</tr>
<tr>
<td>Alpha5</td>
<td>10000</td>
<td>1.2433</td>
<td>0.6977</td>
<td>0.7948</td>
<td>1.2598</td>
<td>1.7255</td>
</tr>
<tr>
<td>Alpha6</td>
<td>10000</td>
<td>-0.8729</td>
<td>0.8040</td>
<td>-1.4033</td>
<td>-0.8692</td>
<td>-0.3276</td>
</tr>
<tr>
<td>Alpha7</td>
<td>10000</td>
<td>-0.9827</td>
<td>0.8346</td>
<td>-1.5247</td>
<td>-0.9646</td>
<td>-0.4223</td>
</tr>
<tr>
<td>Alpha8</td>
<td>10000</td>
<td>0.4771</td>
<td>0.9095</td>
<td>-0.1262</td>
<td>0.4796</td>
<td>1.0952</td>
</tr>
<tr>
<td>TrtX</td>
<td>10000</td>
<td>-1.2319</td>
<td>0.3929</td>
<td>-1.4898</td>
<td>-1.2286</td>
<td>-0.9707</td>
</tr>
<tr>
<td>GenderM</td>
<td>10000</td>
<td>-2.6607</td>
<td>0.5483</td>
<td>-3.0159</td>
<td>-2.6466</td>
<td>-2.2888</td>
</tr>
</tbody>
</table>
Piecewise Exponential Model

Interval estimates:

### Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha1</td>
<td>0.050</td>
<td>-7.4529</td>
<td>-5.5710</td>
</tr>
<tr>
<td>Alpha2</td>
<td>0.050</td>
<td>-5.0961</td>
<td>-3.1973</td>
</tr>
<tr>
<td>Alpha3</td>
<td>0.050</td>
<td>-4.0327</td>
<td>-2.0130</td>
</tr>
<tr>
<td>Alpha4</td>
<td>0.050</td>
<td>-3.1799</td>
<td>-0.7614</td>
</tr>
<tr>
<td>Alpha5</td>
<td>0.050</td>
<td>-0.1893</td>
<td>2.5585</td>
</tr>
<tr>
<td>Alpha6</td>
<td>0.050</td>
<td>-2.4616</td>
<td>0.6875</td>
</tr>
<tr>
<td>Alpha7</td>
<td>0.050</td>
<td>-2.6588</td>
<td>0.6248</td>
</tr>
<tr>
<td>Alpha8</td>
<td>0.050</td>
<td>-1.3264</td>
<td>2.2243</td>
</tr>
<tr>
<td>TrtX</td>
<td>0.050</td>
<td>-2.0147</td>
<td>-0.4735</td>
</tr>
<tr>
<td>GenderM</td>
<td>0.050</td>
<td>-3.7758</td>
<td>-1.6150</td>
</tr>
</tbody>
</table>
### Piecewise Exponential Model

**Hazard ratios of Treatment and Gender:**

\[
\text{hazardratio 'Hazard Ratio Statement 1' Trt;}
\]

\[
\text{hazardratio 'Hazard Ratio Statement 2' Gender;}
\]

#### Bayesian Analysis

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Quantiles</th>
<th>95% Equal-Tail Interval</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt Placebo vs X</td>
<td>10000</td>
<td>3.7058</td>
<td>1.5430</td>
<td>25% 50% 75%</td>
<td>1.6056 7.4981 1.3129 6.7830</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Quantiles</th>
<th>95% Equal-Tail Interval</th>
<th>95% HPD Interval</th>
</tr>
</thead>
</table>
The GENMOD, LIFEREG, and PHREG procedures

LIFEREG: Weibull and exponential models (optional)

Outline

The GENMOD, LIFEREG, and PHREG procedures
- Overview
- Prior distributions
- The BAYES statement
- GENMOD: linear regression
- GENMOD: binomial model
- PHREG: Cox model
- PHREG: piecewise exponential model
- LIFEREG: Weibull and exponential models (optional)
- Summary
**PROC LIFEREG**

- PROC LIFEREG formulates the survival models through an accelerated failure time (AFT) approach:
  \[ y = X \beta + \sigma \epsilon \]

- The response variable \( y \) is usually taken to be the logarithm of the survival time vector \( T \); that is \( y = \log(T) \). \( T \) can be left-, right-, or interval-censored.

- \( \beta \) is a regression parameter vector, \( X \) is a covariates matrix, \( \sigma \) is a dispersion parameter. The \( \epsilon \) parameter represents an error distribution, which leads to different models.

- Model parameters are \((\beta, \sigma)\).
The GENMOD, LIFEREG, and PHREG procedures

LIFEREG: Weibull and exponential models (optional)

PROC LIFEREG

Suppose that $y = \log(T)$ and $\epsilon$ is an extreme value distribution. The resulting model for $y$ is a Weibull distribution. For the Weibull regression model, the density for $t_i$ is

$$f(t_i) = \frac{1}{\sigma} \cdot t_i^{\frac{1}{\sigma} - 1} \exp\left(-\frac{\mu_i}{\sigma}\right) \exp\left\{-t_i^{\frac{1}{\sigma}} \exp\left(-\frac{\mu_i}{\sigma}\right)\right\}$$

where

$$\mu_i = x_i' \beta$$

The exponential regression model is a special case of the Weibull model, with $\sigma = 1$. Hence the density for the survival time is

$$f(t_i) = \exp(-x_i' \beta) \exp\left\{-t_i \exp(-x_i' \beta)\right\}$$
PROC LIFEREG

Suppose that all the responses are observed. The log likelihood is

\[ L = \sum \log \left( \frac{f(u_i)}{\sigma} \right) \]

where

\[ u_i = \frac{(y_i - x_i' \beta)}{\sigma} \]

and \( f(u_i) \) is the density for \( \epsilon_i \).
PROC LIFEREG

If some of the responses are left-, right-, or interval-censored, the log likelihood can be written as

\[ L = \sum \log \left( \frac{f(u_i)}{\sigma} \right) + \sum \log (S(u_i)) \\
+ \sum \log (F(u_i)) + \sum \log (F(u_i) - F(v_i)) \]

with sums over uncensored, right-censored, left-censored, and interval-censored observations, respectively. \( F(u_i) \) is the cdf of \( \epsilon_i \), \( S(u_i) = 1 - F(u_i) \), the corresponding survival function. And \( v_i \) is defined as

\[ v_i = \frac{(z_i - x'_i \beta)}{\sigma} \]

where \( z_i \) is the lower end of a censoring interval.
PROC LIFEREG

Consider the E1690 melanoma clinical trial with \( n = 427 \) subjects.

Fit a parametric survival model with three covariates: Treatment (\( \text{trt} \)), age, and sex. Both \( \text{trt} \) and \( \text{sex} \) are class variables.

The variable \( \text{rfstime} \) is the failure time, and \( \text{rfscens} \) is the censored time.

Two models are considered: Weibull model and exponential model.

A noninformative prior is used on \( \beta \), and a gamma prior is used on the scale parameter.
PROC LIFEREG

Fit Weibull and exponential models to the data

```
proc lifereg data=e1690;
  class trt sex;
  model  rfstime*rfscens(0) = trt age sex / dist=weibull;
  bayes seed=1 outpost=wout;
run;

proc lifereg data=e1690;
  class trt sex;
  model  rfstime*rfscens(0) = trt age sex / dist=exponential;
  bayes seed=1 outpost=eout;
run;
```
PROC LIFEREG

Parts of PROC LIFEREG output:

Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Constant</td>
</tr>
<tr>
<td>trt1</td>
<td>Constant</td>
</tr>
<tr>
<td>age</td>
<td>Constant</td>
</tr>
<tr>
<td>sex1</td>
<td>Constant</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>Gamma</td>
<td>Shape 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inverse Scale 0.001</td>
</tr>
</tbody>
</table>

Independent Prior Distributions for Model Parameters
## Posterior Statistics from the Weibull Model

### Bayesian Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles 25%</th>
<th>Percentiles 50%</th>
<th>Percentiles 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10000</td>
<td>2.5854</td>
<td>0.3874</td>
<td>2.3240</td>
<td>2.5748</td>
<td>2.8379</td>
</tr>
<tr>
<td>trt1</td>
<td>10000</td>
<td>-0.3277</td>
<td>0.1936</td>
<td>-0.4575</td>
<td>-0.3273</td>
<td>-0.1970</td>
</tr>
<tr>
<td>age</td>
<td>10000</td>
<td>-0.0148</td>
<td>0.00736</td>
<td>-0.0196</td>
<td>-0.0146</td>
<td>-0.00988</td>
</tr>
<tr>
<td>sex1</td>
<td>10000</td>
<td>-0.2785</td>
<td>0.2016</td>
<td>-0.4140</td>
<td>-0.2746</td>
<td>-0.1429</td>
</tr>
<tr>
<td>Scale</td>
<td>10000</td>
<td>1.4785</td>
<td>0.0843</td>
<td>1.4199</td>
<td>1.4745</td>
<td>1.5332</td>
</tr>
</tbody>
</table>

### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>1.8596</td>
<td>3.3648</td>
</tr>
<tr>
<td>trt1</td>
<td>0.050</td>
<td>-0.7120</td>
<td>0.0573</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.0296</td>
<td>-0.00058</td>
</tr>
<tr>
<td>sex1</td>
<td>0.050</td>
<td>-0.6786</td>
<td>0.1161</td>
</tr>
<tr>
<td>Scale</td>
<td>0.050</td>
<td>1.3253</td>
<td>1.6535</td>
</tr>
</tbody>
</table>
### Posterior Statistics from the Exponential Model

**Bayesian Analysis**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>Intercept</td>
<td>10000</td>
<td>2.2062</td>
<td>0.2609</td>
<td>2.0299</td>
</tr>
<tr>
<td>trt1</td>
<td>10000</td>
<td>-0.2340</td>
<td>0.1285</td>
<td>-0.3208</td>
</tr>
<tr>
<td>age</td>
<td>10000</td>
<td>-0.0115</td>
<td>0.00499</td>
<td>-0.0149</td>
</tr>
<tr>
<td>sex1</td>
<td>10000</td>
<td>-0.2213</td>
<td>0.1357</td>
<td>-0.3122</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>1.6961</td>
<td>2.7160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.6791</td>
</tr>
<tr>
<td>trt1</td>
<td>0.050</td>
<td>-0.4854</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.4800</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.0213</td>
<td>-0.00162</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.0211</td>
</tr>
<tr>
<td>sex1</td>
<td>0.050</td>
<td>-0.4907</td>
<td>0.0405</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.4799</td>
</tr>
</tbody>
</table>
Kernel Density Comparison Plots of the Parameters

- **Intercept**
- **Treatment**
- **Age**
- **Sex**

**Weibull Model** vs. **Exponential Model**
PROC LIFEREG

Which model fits the data better? You can use the DIC as a model selection criterion. In this case, the Weibull model (top table) gives a much smaller DIC, and it gives the better fit to the data than the exponential model.

Bayesian Analysis

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DIC (smaller is better)</td>
<td>1302.220</td>
</tr>
<tr>
<td>pD (effective number of parameters)</td>
<td>4.926</td>
</tr>
</tbody>
</table>

Bayesian Analysis

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DIC (smaller is better)</td>
<td>1352.719</td>
</tr>
<tr>
<td>pD (effective number of parameters)</td>
<td>3.920</td>
</tr>
</tbody>
</table>
Summary

- SAS has developed capabilities for Bayesian analysis in its existing generalized linear models and survival analysis software.
- SAS is in the process of adding more features to these procedures. Currently, the following procedure statements are ignored when the BAYES statement is specified:

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Ignored Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENMOD</td>
<td>REPEATED, ASSESS, CONTRAST, ESTIMATE, OUTPUT, LSMEANS</td>
</tr>
<tr>
<td>LIFEREG</td>
<td>OUTPUT, PROBPLOT</td>
</tr>
<tr>
<td>PHREG</td>
<td>ACCESS, CONTRAST, OUTPUT, TEST</td>
</tr>
</tbody>
</table>

- See the SAS/STAT® User’s Guide for more information.
Experimental Downloads for SAS® 9.1.3

In SAS 9.1.3, you can download BGENMOD, BLIFEREG, and BPHREG procedures that have the Bayesian capabilities.

In SAS 9.1.3,

- No BY-group processing
- No OUTPOST= option. To save the posterior samples, use the ODS statement:
  `ODS OUTPUT PosteriorSample = SAS-data-set;`
- DIC available only in PROC BPHREG
- BPHREG does not support the BASELINE statement, and does not have the HAZARDRATIO statement.
References

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.
Bayesian Modeling Using the MCMC Procedure

Joseph G. Ibrahim

Department of Biostatistics

University of North Carolina
Outline

1. Overview
2. Getting started: linear regression
3. How does PROC MCMC work?
4. Essential statements
5. Distributions
6. More examples
7. Closing remarks
Bayesian Computation in SAS® Software

- The GENMOD, LIFEREG, and PHREG procedures:
  - use well defined models
  - provide some choices of prior distributions
  - require a minimal amount of programming
  - became production in SAS 9.2 (the BAYES statement)

- The MCMC procedure is:
  - flexible and capable of handling more general models
  - programming-oriented
  - experimental in SAS 9.2 and will become production during 2009
The MCMC procedure is a general-purpose simulation procedure that uses Markov chain Monte Carlo (MCMC) techniques to fit a range of Bayesian models:

- single-level or multilevel (hierarchical) models
- linear or nonlinear models, such as regression, mixture, survival, ordinal multinomial, and so on.
The MCMC Procedure

The MCMC procedure has the following characteristics:

- **Ease-of-use**: You have to specify only parameters, prior distributions, and a likelihood function.

- **Generality**: You can analyze data that have standard distribution (normal, gamma, binomial) or general likelihood or prior distribution, as long as they are programmable using the SAS DATA step functions.
Posterior Statistics

In addition to samples from the desired posterior distribution, PROC MCMC produces:

- **posterior statistics**:  
  - posterior mean, standard deviation, percentiles  
  - equal-tail and highest posterior density (HPD) intervals  
  - covariance/correlation matrices  
  - deviance information criterion (DIC)

- **Markov chain convergence diagnostics**:  
  - Geweke test  
  - Heidelberger-Welch stationarity and half-width tests  
  - Raftery-Lewis test  
  - posterior sample autocorrelations  
  - effective sample size (ESS)  
  - Monte Carlo standard error (MCSE)
Visualization

Graphical display of the posterior samples:

- trace plot (with optional smoothed mean curve)
- autocorrelation plot
- kernel density plot (with optional fringe plot)
Outline

1. Overview

2. Getting started: linear regression

3. How does PROC MCMC work?

4. Essential statements

5. Distributions

6. More examples

7. Closing remarks
Linear Regression

Consider the model

\[ \text{Weight}_i = \beta_0 + \beta_1 \text{Height}_i + \epsilon_i \]

for the observations \( i = 1, \cdots, n \) and \( \epsilon \) is a \( N(0, \sigma^2) \) error term.

The following DATA step creates data set Class with variables Height and Weight:

data Class;
  input Name $ Height Weight @@;
datalines;
Alfred 69.0 112.5  Alice 56.5 84.0  Barbara 65.3 98.0
Carol 62.8 102.5  Henry 63.5 102.5  James 57.3 83.0
Jane 59.8 84.5  Janet 62.5 112.5  Jeffrey 62.5 84.0
John 59.0 99.5  Joyce 51.3 50.5  Judy 64.3 90.0
Louise 56.3 77.0  Mary 66.5 112.0  Philip 72.0 150.0
Robert 64.8 128.0  Ronald 67.0 133.0  Thomas 57.5 85.0
William 66.5 112.0
;
Linear Regression

The likelihood function is a normal density:

$$\text{Weight}_i \sim \text{normal}(\beta_0 + \beta_1 \text{Height}_i, \sigma^2)$$

Suppose that you want to consider the following prior distributions:

$$\pi(\beta_0) = \phi(0, \text{var} = 1e6)$$
$$\pi(\beta_1) = \phi(0, \text{var} = 1e6)$$
$$\pi(\sigma^2) = f_{\Gamma}(\text{shape} = 3/10, \text{scale} = 10/3)$$
Linear Regression

The following statements fit a simple linear regression:

```
proc mcmc data=class outpost=classout nmc=50000 thin=5 seed=246810;
  parms beta0 0 beta1 0;
  parms sigma2 1;
  prior beta0 beta1 ~ normal(mean=0, var=1e6);
  prior sigma2 ~ igamma(shape=3/10, scale=10/3);
  mu = beta0 + beta1*height;
  model weight ~ normal(mu, var=sigma2);
run;
```
Diagnostics Plots for $\beta_0$
Diagnostics Plots for $\beta_1$
Diagnostics Plots for $\sigma^2$
### Observation and Prior Information Tables

<table>
<thead>
<tr>
<th>Block</th>
<th>Parameter</th>
<th>Sampling Method</th>
<th>Initial Value</th>
<th>Prior Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>beta0</td>
<td>N-Metropolis</td>
<td>0</td>
<td>normal(mean = 0, var = 1e6)</td>
</tr>
<tr>
<td>1</td>
<td>beta1</td>
<td>N-Metropolis</td>
<td>0</td>
<td>normal(mean = 0, var = 1e6)</td>
</tr>
<tr>
<td>2</td>
<td>sigma2</td>
<td>N-Metropolis</td>
<td>1.0000</td>
<td>igamma(shape = 3/10, scale = 10/3)</td>
</tr>
</tbody>
</table>

Number of Observations Read: 19
Number of Observations Used: 19
Tuning History Table Monitors the Tuning of the Metropolis Sampler

<table>
<thead>
<tr>
<th>Phase</th>
<th>Block</th>
<th>Scale</th>
<th>Acceptance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2.3800</td>
<td>0.0420</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.3800</td>
<td>0.8860</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.0938</td>
<td>0.2180</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15.5148</td>
<td>0.3720</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.8299</td>
<td>0.4860</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15.5148</td>
<td>0.1260</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.1132</td>
<td>0.4840</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.4767</td>
<td>0.0880</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.4866</td>
<td>0.5420</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.1914</td>
<td>0.2000</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2.2784</td>
<td>0.4600</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.7859</td>
<td>0.3900</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2.8820</td>
<td>0.3360</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.7859</td>
<td>0.4020</td>
</tr>
</tbody>
</table>
## Burn-In and Sampling History

<table>
<thead>
<tr>
<th>Block</th>
<th>Scale</th>
<th>Acceptance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8820</td>
<td>0.3400</td>
</tr>
<tr>
<td>2</td>
<td>3.7859</td>
<td>0.4150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block</th>
<th>Scale</th>
<th>Acceptance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8820</td>
<td>0.3284</td>
</tr>
<tr>
<td>2</td>
<td>3.7859</td>
<td>0.4008</td>
</tr>
</tbody>
</table>
### Summary Statistics and Posterior Intervals Tables

#### Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>10000</td>
<td>-142.6</td>
<td>33.9390</td>
<td>-164.5</td>
<td>-142.4</td>
<td>-120.5</td>
</tr>
<tr>
<td>beta1</td>
<td>10000</td>
<td>3.8917</td>
<td>0.5427</td>
<td>3.5406</td>
<td>3.8906</td>
<td>4.2402</td>
</tr>
<tr>
<td>sigma2</td>
<td>10000</td>
<td>136.8</td>
<td>51.7417</td>
<td>101.8</td>
<td>126.0</td>
<td>159.9</td>
</tr>
</tbody>
</table>

#### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>0.050</td>
<td>-209.3</td>
<td>-209.7</td>
</tr>
<tr>
<td>beta1</td>
<td>0.050</td>
<td>2.8317</td>
<td>2.8280</td>
</tr>
<tr>
<td>sigma2</td>
<td>0.050</td>
<td>69.2208</td>
<td>58.2627</td>
</tr>
</tbody>
</table>
MCSE and Autocorrelation Tables

### Monte Carlo Standard Errors

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MCSE</th>
<th>Standard Deviation</th>
<th>MCSE/SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>0.4576</td>
<td>33.9390</td>
<td>0.0135</td>
</tr>
<tr>
<td>beta1</td>
<td>0.00731</td>
<td>0.5427</td>
<td>0.0135</td>
</tr>
<tr>
<td>sigma2</td>
<td>0.7151</td>
<td>51.7417</td>
<td>0.0138</td>
</tr>
</tbody>
</table>

### Posterior Autocorrelations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lag 1</th>
<th>Lag 5</th>
<th>Lag 10</th>
<th>Lag 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>0.2986</td>
<td>-0.0008</td>
<td>0.0162</td>
<td>0.0193</td>
</tr>
<tr>
<td>beta1</td>
<td>0.2971</td>
<td>0.0000</td>
<td>0.0135</td>
<td>0.0161</td>
</tr>
<tr>
<td>sigma2</td>
<td>0.2966</td>
<td>0.0062</td>
<td>0.0008</td>
<td>-0.0068</td>
</tr>
</tbody>
</table>
## Geweke Diagnostics and ESS Tables

<table>
<thead>
<tr>
<th>Geweke Diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>beta0</td>
</tr>
<tr>
<td>beta1</td>
</tr>
<tr>
<td>sigma2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effective Sample Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>beta0</td>
</tr>
<tr>
<td>beta1</td>
</tr>
<tr>
<td>sigma2</td>
</tr>
</tbody>
</table>
Outline

1. Overview
2. Getting started: linear regression
3. How does PROC MCMC work?
4. Essential statements
5. Distributions
6. More examples
7. Closing remarks
Sampling in PROC MCMC

The basic sampling algorithm that drives PROC MCMC is a normal-kernel-based random walk Metropolis. The proposal distribution is $q(\theta_{\text{new}}|\theta^{(t)}) = \text{MVN}(\theta_{\text{new}}|\theta^{(t)}, c^2 \Sigma)$.

Two crucial components in the Metropolis algorithm:

- construction of the proposal distribution—automatically done by PROC MCMC
- evaluation of $\log(\pi(\theta^{(t)}|y))$ at each iteration—you specify the prior and likelihood function, and PROC MCMC does the rest
Computations in PROC MCMC

At each Markov chain iteration, PROC MCMC calculates \( \log(\pi(\theta^{(t)}|y)) \) by stepping through the input data set. When you use the procedure, keep this equation in mind:

\[
\log(\pi(\theta|y)) = \log(\pi(\theta)) + \sum_{i=1}^{n} \log(f(y_i|\theta)) + C
\]

where \( \theta \) are model parameters, \( \pi(\theta) \) is the prior, \( f(y_i|\theta) \) is the sampling distribution for a single observation in the data set, and \( C \) is a constant that can be ignored.

To calculate \( \log(\pi(\theta^{(t)}|y)) \), PROC MCMC steps through the data set, performs the computations for each \( y_i \), and cumulatively adds the log-likelihood values.
How does PROC MCMC work?

Computations in PROC MCMC

You are not restricted to models that have the same likelihood function for each observation. PROC MCMC enables you to model data that have the following likelihood function:

$$
\log(f(y|\theta)) = \sum_{i=1}^{n_1} \log(f_1(y_i|\theta)) + \sum_{i=n_1+1}^{n} \log(f_2(y_i|\theta))
$$

$$
\log(f(y|\theta)) = \sum_{i=1}^{n} \log(f(y_i|y_{j\neq i}, \theta))
$$

At each simulation step, PROC MCMC processes the data set, evaluates the program, and calculates the log of the posterior density.
Outline

4 Essential statements
- Basic statements
- PROC statement and options
- PARMS statement
- PRIOR statement
- MODEL statement
- Programming statements
PROC MCMC options;

PARMS; define parameters.

BEGINCNST; set up statements, evaluated before but not during the simulation
Programming statements;
ENDCNST;

BEGINPRIOR; calculate \( \log(\pi(\theta)) \) or \( g(\theta) \) (statements should not contain data set variables)
Programming statements;
ENDPRIOR;

PRIOR declare prior distributions
PRIOR

Programming statements; \( \log(f(y_i|\theta)) \), evaluated \( n \)-times per iteration

MODEL

Run;
**PROC MCMC options**

The PROC MCMC statement invokes the procedure. Some useful options are:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA=</td>
<td>name of the input data set</td>
</tr>
<tr>
<td>OUTPOST=</td>
<td>name of the output data set for posterior samples</td>
</tr>
<tr>
<td>NBI=</td>
<td>number of burn-in iterations</td>
</tr>
<tr>
<td>NMC=</td>
<td>number of MCMC iterations</td>
</tr>
<tr>
<td>THIN=</td>
<td>thinning of the Markov chain</td>
</tr>
<tr>
<td>SEED=</td>
<td>random number generator seed</td>
</tr>
<tr>
<td>STATISTICS=</td>
<td>posterior statistics</td>
</tr>
<tr>
<td>DIAGNOSTICS=</td>
<td>convergence diagnostics</td>
</tr>
<tr>
<td>PLOTS=</td>
<td>diagnostics plotting</td>
</tr>
<tr>
<td>MONITOR=</td>
<td>analysis for selected symbols of interest, (for example, functions of parameters)</td>
</tr>
</tbody>
</table>
**PARMS Statement**

**PARMS** \( name \mid (name\text{-}list) \leq number; \)

lists the names of the parameters and specifies optional initial values. PROC MCMC generates values for uninitialized parameters from the corresponding prior distributions.

For example:

```
PARMS alpha 0 beta 1;
```

declares \( \alpha \) and \( \beta \) to be model parameters and assigns 0 to \( \alpha \) and 1 to \( \beta \).

```
PARMS alpha 0 beta;
```

assigns 0 to \( \alpha \) and leaves \( \beta \) uninitialized.

```
PARMS (alpha beta) 1;
```

assigns 1 to both \( \alpha \) and \( \beta \).
Initial Values

You can use the PARMS statement to assign initial values to model parameters. If the initial values lead to an invalid prior or likelihood calculation, PROC MCMC prints an error message and stops.

For example,

```
parm sigma2 -1;
prior sigma2 ~ igamma(shape = 3/10, scale = 10/3);
```

leads to the following error message:

```
ERROR: The initial value -1 for parameter sigma2 is outside of the prior distribution support set.
NOTE: The prior of sigma2 is GAMMA with SHAPE=0.01. It has a support set of (0, Infinity).
```

In some cases, PROC MCMC provides additional information.
PARMS Statement

When multiple PARMS statements are used, each statement defines a block of parameters. PROC MCMC updates parameters in each block sequentially, conditional on the current values of other parameters in other blocks. For example, in the linear regression example, the following PARMS statements are used:

```
PARMS beta0 beta1;
PARMS sigma2;
```

At each iteration $t$, PROC MCMC updates $\beta_0$ and $\beta_1$ together, alternatively with $\sigma^2$, each with a Metropolis sampler:

$$
\begin{align*}
\beta_0^{(t)}, \beta_1^{(t)} & \mid \sigma^2_{(t-1)} \\
\sigma^2_{(t)} & \mid \beta_0^{(t)}, \beta_1^{(t)}
\end{align*}
$$
## Prior Information Table

Blocking information is included in the table.

<table>
<thead>
<tr>
<th>Block</th>
<th>Parameter</th>
<th>Sampling Method</th>
<th>Initial Value</th>
<th>Prior Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>beta0</td>
<td>N-Metropolis</td>
<td>0</td>
<td>normal(mean = 0, var = 1e6)</td>
</tr>
<tr>
<td>1</td>
<td>beta1</td>
<td>N-Metropolis</td>
<td>0</td>
<td>normal(mean = 0, var = 1e6)</td>
</tr>
<tr>
<td>2</td>
<td>sigma2</td>
<td>N-Metropolis</td>
<td>1.0000</td>
<td>igamma(shape = 3/10, scale = 10/3)</td>
</tr>
</tbody>
</table>
PRIOR Statement

PRIOR parameter-list ~ distribution;

specifies the prior distributions. The parameter-list can be a single parameter or a list of parameters. Multiple PRIOR statements are allowed and you can have as many hierarchical levels as desired. For example:

PRIOR alpha ~ normal(0, var=10);
PRIOR sigma2 ~ igamma(0.001, iscale=0.001);
PRIOR beta gamma ~ normal(alpha, var=sigma2);

specifies the following joint prior distribution:

\[ \pi(\alpha, \beta, \gamma, \sigma^2) = \pi(\beta|\alpha, \sigma^2) \cdot \pi(\gamma|\alpha, \sigma^2) \cdot \pi(\alpha) \cdot \pi(\sigma^2) \]

PROC MCMC includes a HYPER statement, which is treated the same as the PRIOR statement. It is a notational convenience for specifying a multilevel hierarchical model.
**MODEL Statement**

**MODEL** dependent-variable-list $\sim$ distribution;

specifies the conditional distribution of the data given the parameters (the likelihood function). The dependent variables can be either variables from the data set or functions of variables in the program. If $y$ is a variable in the data set,

```
MODEL y $\sim$ normal(alpha, var=1);
```

specifies

$$\pi(y_i \mid \alpha) = \phi(y_i \mid \alpha, 1)$$

If $w = \log(y)$,

```
w = log(y);
MODEL w $\sim$ normal(alpha, var=1);
```

specifies

$$\pi(\log(y_i) \mid \alpha) = \phi(\log(y_i) \mid \alpha, 1)$$

Multiple MODEL statements are allowed.
Programming Statements

Most DATA step operators, functions, and statements can be used in PROC MCMC:

- assignment and operators: +, -, *, /, <>, <, ...
- mathematical functions: PDF, CDF, SDF, LOGPDF, ABS, LOG, LOGISTIC, FACT, BETA, GAMMA, RAND, ...
- statements: CALL, DO, IF, PUT, WHEN, ...

The functions enable you to:

- construct log of density functions (both priors and likelihood functions)
- compute functions of parameters
- generate samples from the predictive distribution
- debug your program

PROC MCMC also supports matrix-based functions.
The programming order matters. For example, the following linear regression statements will not work correctly because \( \mu \) is defined after the MODEL statement:

```sas
proc mcmc data=class outpost=classout nmc=50000 thin=5
  seed=246810;
  parms beta0 0 beta1 0;
  parms sigma2 1;
  prior beta0 beta1 ~ normal(mean=0, var=1e6);
  prior sigma2 ~ igamma(shape=3/10, scale=10/3);
  model weight ~ normal(mu, var=sigma2);
  mu = beta0 + beta1*height;
run;
```
Outline

5 Distributions
- Standard distributions
- A list of standard distributions
- Binomial model
- Nonlinear Poisson regression
- Change point model
- Nonstandard distributions
- The GENERAL function
- Linear regression with improper prior
- Zero-inflated Poisson regression
Standard Distributions

PROC MCMC supports the following standard distributions that can be used in both the PRIOR and MODEL statements:

<table>
<thead>
<tr>
<th>beta</th>
<th>binary</th>
<th>binomial</th>
<th>cauchy</th>
<th>chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>expon</td>
<td>gamma</td>
<td>geo</td>
<td>ichisq</td>
<td>igamma</td>
</tr>
<tr>
<td>laplace</td>
<td>negbin</td>
<td>normal</td>
<td>pareto</td>
<td>poisson</td>
</tr>
<tr>
<td>sichisq</td>
<td>t</td>
<td>uniform</td>
<td>wald</td>
<td>weibull</td>
</tr>
</tbody>
</table>

Distribution argument can be constants, expressions, or model parameters. For example:

```
prior alpha ~ cauchy(0, 2);
prior p ~ beta(abs(alpha), constant('pi'));
model y ~ binomial(n, p);
```
Standard Distributions

Some distributions can be parameterized in different ways:

<table>
<thead>
<tr>
<th>Standard Distributions</th>
<th>Parameterized Ways</th>
</tr>
</thead>
<tbody>
<tr>
<td>expon(scale</td>
<td>s = λ)</td>
</tr>
<tr>
<td>gamma(a, scale</td>
<td>sc = λ)</td>
</tr>
<tr>
<td>igamma(a, scale</td>
<td>sc = λ)</td>
</tr>
<tr>
<td>laplace(l, scale</td>
<td>sc = λ)</td>
</tr>
<tr>
<td>normal(μ, var = σ²)</td>
<td>normal(μ, sd = σ)</td>
</tr>
<tr>
<td>lognormal(μ, var = σ²)</td>
<td>lognormal(μ, sd = σ)</td>
</tr>
<tr>
<td>t(μ, var = σ², df)</td>
<td>t(μ, sd = σ, df)</td>
</tr>
</tbody>
</table>

For these distributions, you must explicitly name the ambiguous parameter. For example:

```plaintext
prior beta ~ normal(0, var=sigma2);
prior sigma2 ~ igamma(0.001, is=0.001);
```
Binomial Model

Researchers are interested in evaluating the performance of a medical procedure in a multicenter study. One of the study goals is to compare the survival benefit of the medical procedure. The following statements create a SAS data set for the treatment arm of the trials:

```sas
data trials;
  input event n center;
datalines;
  2 86 1
  2 69 2
  1 71 3
  1 113 4
  1 103 5
;
```

- **event**: number of deaths
- **n**: number of patients assigned to the treatment procedure
- **center**: center index
Binomial Model

The simplest model ignores any center difference and treats the data as the realization of a shared model, with the same death probability \( p \) applied to all centers:

\[
\text{event}_i \sim \text{binomial}(n_i, p) \\
\pi(p) = \text{uniform}(0, 1)
\]
Binomial Model

Fitting a binomial model in PROC MCMC with a uniform prior on $p$ is straightforward:

```sas
proc mcmc data=trials seed=17;
  parm p;
  prior p ~ beta(1,1);
  model event ~ binomial(n,p);
run;
```
### Summary Statistics and Posterior Intervals Tables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>p</td>
<td>10000</td>
<td>0.0180</td>
<td>0.00629</td>
<td>0.0134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.050</td>
<td>0.00795</td>
<td>0.0321</td>
</tr>
</tbody>
</table>
Binomial Model

Suppose that you do not want to have fixed hyperparameter values and want to consider hyperprior distributions on these parameters:

\[ \pi(a) \propto \text{exponential}(\text{scale} = 100) \]
\[ \pi(b) \propto \text{exponential}(\text{scale} = 100) \]

This prior has mean of 100 and variance 10,000. The following SAS statements fit a hierarchical binomial model:

```sas
proc mcmc data=trials seed=17 nmc=10000 outpost=bmc;
   parms p;
   parms a b;
   hyper a b ~ expon(scale=100);
   prior p ~ beta(a,b);
   model event ~ binomial(n,p);
run;
```
Posterior Density Comparison

Having hyperprior distributions is essentially equivalent to using a uniform prior on $p$—there is no information in the data that can help with estimating the hyperparameters.
Truncated Distributions

Most standard distributions, with the exception of the binary and uniform, allow for optional LOWER= and UPPER= arguments.

For example, in the binomial model, if you believe that the success probability cannot be above a certain threshold, in addition to restrictions on the beta parameters, you can use the following statements:

```
proc mcmc data=trials seed=17 nmc=10000 outpost=bmc;
  parms p;
  parms a b;
  hyper a ~ expon(scale=100, lower=45);
  hyper b ~ expon(scale=100, lower=100, upper=2000);
  prior p ~ beta(a, b, upper=0.1);
  model event ~ binomial(n,p);
run;
```
Nonlinear Poisson Regression

This example analyzes calls to a technical center after a product release. The information can be used to determine the allocation of technical support resources for future products.

data calls;
    input weeks calls @@;
    datalines;
1 0 2 2 3 1 4 8
5 5 6 17 7 24 8 23
9 19 10 17
;

The variable weeks is the number of weeks. The variable calls counts the number of calls.
Nonlinear Poisson Regression

You can model calls as a Poisson random variable, with the mean modeled as a nonlinear function of weeks:

\[
\text{calls}_i \sim \text{Poisson} (\lambda_i)
\]

The mean function \( \lambda_i \) is modeled as:

\[
\lambda_i = \frac{\gamma}{1 + \exp[-(\alpha + \beta \text{weeks}_i)]}
\]

The prior distributions are:

\[
\pi(\alpha) \sim \text{normal}(-5, \text{sd} = .25)
\]
\[
\pi(\beta) \sim \text{normal}(0.75, \text{sd} = .5)
\]
\[
\pi(\gamma) \sim \text{gamma}(3.5, \text{scale} = 12)
\]
Nonlinear Poisson Regression

The following statements fit a nonlinear Poisson regression to the calls data.

```sas
proc mcmc data=calls nmc=20000 propcov=quanew;
  parms alpha beta gamma;
  prior alpha ~ normal(-5, sd=0.25);
  prior beta ~ normal(0.75, sd=0.5);
  prior gamma ~ gamma(3.5, scale=12);
  lambda = gamma*logistic(alpha+beta*weeks);
  model calls ~ poisson(lambda);
run;
```

The PROPCOV= option obtains optimal starting values and constructs the proposal distribution via numerical optimization.
Nonlinear Poisson Regression

![Graph of Nonlinear Poisson Regression](image-url)
Change Point Model

Consider the following data:

<table>
<thead>
<tr>
<th>OBS</th>
<th>y</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.12</td>
<td>-1.39</td>
</tr>
<tr>
<td>2</td>
<td>1.12</td>
<td>-1.39</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>-1.08</td>
</tr>
<tr>
<td>4</td>
<td>1.03</td>
<td>-1.08</td>
</tr>
<tr>
<td>5</td>
<td>0.92</td>
<td>-0.94</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Change Point Model

Let the change point be $cp$. You can use the regression model

$$y_i \sim \begin{cases} 
\text{normal}(\alpha + \beta_1(x_i - cp), \sigma^2) & \text{if } x_i < cp \\
\text{normal}(\alpha + \beta_2(x_i - cp), \sigma^2) & \text{if } x_i \geq cp 
\end{cases}$$

with the following prior distributions:

$$\pi(cp) \propto \text{uniform}(-1.3, 1.1)$$
$$\pi(\alpha, \beta_1, \beta_2) \propto \text{normal}(0, \text{var} = 1e6)$$
$$\pi(\sigma) \propto \text{uniform}(0, 10)$$
IF/ELSE Logical Control

```sas
proc mcmc data=stagnant outpost=outc propcov=quanew
   seed=23 nmc=20000;

   parms alpha cp beta1 beta2;
   parms s2;

   prior cp ~ unif(-1.3, 1.1);
   prior alpha beta: ~ normal(0, v = 1e6);
   prior s2 ~ uniform(0, 10);

   if(x < cp) then
      mu = alpha + beta1 * (x - cp);
   else
      mu = alpha + beta2 * (x - cp);
   model y ~ normal(mu, sig=s2);
run;
```
Change Point Model

Posterior estimates of the regression fit and the posterior marginal distribution of the change point location:
But Wait!

Does it have to be a change point model? How about a quadratic regression?
Quadratic Regression

Fitting a quadratic regression is straightforward:

```plaintext
proc mcmc data=stagnant outpost=outc propcov=quanew
   seed=23 nmc=20000;

   parms alpha beta1 beta2;
   parms s2;

   prior alpha beta: ~ normal(0, v = 1e6);
   prior s2 ~ uniform(0, 10);
   mu = alpha + beta1 * x + beta2 * x * x;
   model y ~ normal(mu, var=s2);
run;
```
The Fit Doesn’t Look Bad!

Posterior estimates of the quadratic regression fit:

How can you decide which model to use?
Use Deviance Information Criterion!
To request DIC:

```sas
proc mcmc data=stagnant outpost=outq propcov=quanew
seed=23 nmc=20000 dic;
```

<table>
<thead>
<tr>
<th>Change Point Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance Information Criterion</td>
</tr>
<tr>
<td>Dbar (posterior mean of deviance)</td>
</tr>
<tr>
<td>Dmean (deviance evaluated at posterior mean)</td>
</tr>
<tr>
<td>pD (effective number of parameters)</td>
</tr>
<tr>
<td>DIC (smaller is better)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadratic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance Information Criterion</td>
</tr>
<tr>
<td>Dbar (posterior mean of deviance)</td>
</tr>
<tr>
<td>Dmean (deviance evaluated at posterior mean)</td>
</tr>
<tr>
<td>pD (effective number of parameters)</td>
</tr>
<tr>
<td>DIC (smaller is better)</td>
</tr>
</tbody>
</table>
Outline

5 Distributions

- Standard distributions
- A list of standard distributions
- Binomial model
- Nonlinear Poisson regression
- Change point model
- Nonstandard distributions

The GENERAL function

- Linear regression with improper prior
- Zero-inflated Poisson regression
Specifying a New Distribution

The GENERAL and DGENERAL functions enable you to analyze data that have any prior or likelihood functions, as long as these functions are programmable using the SAS DATA step functions. The “D” stands for discrete.

PRIOR alpha ~ dgeneral(lp);
MODEL y ~ general(llike);

The expressions lp and llike must take the values of the logarithm of the prior density or likelihood function that you construct using SAS programming statements.

The normalizing constant of the distribution can be ignored, as long as it is independent of other parameters in the model.
More on the GENERAL Functions

The function argument can be an expression or a constant. For example, to specify an improper flat prior on the real axis, $\pi(\alpha) \propto 1$, you use the following statement:

```
prior alpha ~ general(0);
```

You should be careful using these functions because PROC MCMC cannot verify that the priors you specify are valid (integrable) distributions, and you can easily construct prior and log-likelihood functions that lead to improper posterior distributions.

Use the DGENERAL function if the parameters take only discrete values. PROC MCMC returns continuous values otherwise.
Linear Regression

Suppose that in the “Getting Started” linear regression example, you want to use a noninformative prior on the variance parameter

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2}$$

which is a nonstandard distribution (nonintegrable prior). The logarithm of this prior is

$$\log(\pi(\sigma^2)) = -\log(\sigma^2) + C$$

Note that the normalizing constant can be ignored in PROC MCMC.
Linear Regression

The following statements fit a simple linear regression with noninformative prior on $\sigma^2$:

```plaintext
proc mcmc data=class outpost=classout nmc=50000 thin=5
   seed=246810;
   parms beta0 0 beta1 0;
   parms sigma2 1;
   prior beta0 beta1 ~ normal(mean=0, var=1e6);
   prior sigma2 ~ general(-log(sigma2));
   mu = beta0 + beta1*height;
   model weight ~ normal(mu, var=sigma2);
run;
```
Linear Regression

The argument to the GENERAL function can be any expression:

```
proc mcmc data=class outpost=classout nmc=50000 thin=5
  seed=246810;
  parms beta0 0 beta1 0;
  parms sigma2 1;
  prior beta0 beta1 ~ normal(mean=0, var=1e6);
  lp = -log(sigma2);
  prior sigma2 ~ general(lp);
  mu = beta0 + beta1*height;
  model weight ~ normal(mu, var=sigma2);
run;
```
Zero-Inflated Poisson (ZIP) Regression

You can use the ZIP models when you observe data that have a large number of zeros. A large number of zeros suggests that you might have two populations in the data set, and you typically use a mixture of two distributions to model the data:

$$\Pr(Y = y) = \eta p_1 + (1 - \eta) p_2(y | \mu)$$

where

$$p_1 = \begin{cases} 1 & \text{if } y = 0 \\ 0 & \text{if } y \neq 0 \end{cases}$$

$$p_2(y | \mu) = \text{Poisson}(\mu)$$

$$0 \leq \eta \leq 1$$
ZIP Model

The following data are part of a hypothetical data set ($n = 52$), which represents the number of fish caught by visitors at a state park. Variables are the visitor’s age and gender.

<table>
<thead>
<tr>
<th>OBS</th>
<th>age</th>
<th>count</th>
<th>female</th>
<th>male</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>18</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>48</td>
<td>12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

...
Visualize the Data

The top three panels are the histograms for fish caught by all, females, and males. The bottom plots are scatter plots of age versus fish caught (all, females, and males). The curves are penalized B-spline fit.
ZIP Model

To model a mixture of two populations in the data, consider the model

\[
count_i \sim \eta p_{1i} + (1 - \eta)\text{Poisson}(\mu_i)
\]

\[
\log(\mu_i) = \beta_0 + \beta_1 \cdot \text{female}_i \cdot \text{age}_i + \beta_2 \cdot \text{male}_i \cdot \text{age}_i
\]

with priors

\[
\pi(\beta_0, \beta_1, \beta_2) = \text{normal}(0, \sigma^2 = 1000)
\]

\[
\pi(\eta) = \text{uniform}(0, 1)
\]

You can use the DGENERAL function to specify the mixture likelihood function.
ZIP Model

The following SAS statements fit a ZIP model:

```sas
proc mcmc data=catch seed=17 nmc=10000
   propcov=quanew plots=density;
parms beta0 0 beta1 0 beta2 0;
parms eta .3;
prior beta: ~ normal(0,var=1000);
prior eta ~ uniform(0,1);
mu = exp(beta0 + beta1*female*age + beta2*male*age);
llike=log(eta*(count eq 0) + (1-eta)*pdf("poisson",count,mu));
model dgeneral(llike);
run;
```

This type of mixture prior is similar to the spike-and-slab (or lump-and-smear) prior.
Posterior Estimates of the ZIP Model

### Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>beta0</td>
<td>10000</td>
<td>-3.5569</td>
<td>0.6498</td>
<td>-3.9770</td>
</tr>
<tr>
<td>beta1</td>
<td>10000</td>
<td>0.1222</td>
<td>0.0135</td>
<td>0.1127</td>
</tr>
<tr>
<td>beta2</td>
<td>10000</td>
<td>0.1057</td>
<td>0.0140</td>
<td>0.0958</td>
</tr>
<tr>
<td>eta</td>
<td>10000</td>
<td>0.3067</td>
<td>0.0942</td>
<td>0.2396</td>
</tr>
</tbody>
</table>

### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>0.050</td>
<td>-4.8720 -2.3380</td>
<td>-4.7159 -2.1958</td>
</tr>
<tr>
<td>beta1</td>
<td>0.050</td>
<td>0.0970 0.1493</td>
<td>0.0975 0.1496</td>
</tr>
<tr>
<td>beta2</td>
<td>0.050</td>
<td>0.0792 0.1341</td>
<td>0.0797 0.1343</td>
</tr>
<tr>
<td>eta</td>
<td>0.050</td>
<td>0.1319 0.4938</td>
<td>0.1225 0.4808</td>
</tr>
</tbody>
</table>
Posterior Marginal Density Plots

![Posterior Density Plots](image)

- **beta0**: Distribution of the intercept.
- **beta1**: Distribution of the first coefficient.
- **beta2**: Distribution of the second coefficient.
- **eta**: Distribution of the offset.
ZIP Model Fit

Bayesian fit (dotted lines) of the ZIP model. The other two lines are penalized B-spline fit.
Outline

6 More examples
- Functions of parameters
- Posterior predictive distribution
- Incorporation of historical data (power prior)
- Sensitivity analysis
- Random-effects model
Inference on Functions of Parameters

One advantage of Bayesian inference is its ability to estimate functions of the model parameters, such as any $f(\theta)$. In PROC MCMC, it is convenient to obtain samples from the posterior marginal distributions of the unknown of interest. You use:

- programming statements to calculate functions of parameters
- MONITOR= option to select a list of symbols to make inference on
ZIP Model

In the ZIP model, suppose that you want to ask:

- Is the difference between $\beta_1$ and $\beta_2$ significant? That is, how strong a claim can you make that females tend to catch more fish than males, when controlled by age?

- How many more fish is a female expected to catch than a male with an increase of 10 years in age?

These questions can be answered through the examinations of functions of the model parameters, which are random variables and hence have their own posterior distributions:

- $\pi (\beta_1 - \beta_2 > 0 | \text{Data})$
- $\pi (\exp ((\beta_1 - \beta_2) \times 10) | \text{Data})$
Monitoring Functions of Parameters

The MONITOR= options enables you to monitor any expression you created in the program (in addition to the model parameters):

```sas
proc mcmc data=catch seed=17 nmc=10000
   propcov=quanew plots=density monitor=(bdif mdif);
   parms beta0 0 beta1 0 beta2 0;
   parms eta .3;
   prior beta: ~ normal(0,var=1000);
   prior eta ~ uniform(0,1);
   bdif = beta1 - beta2;
   mdif = exp(bdif * 10);
   mu = exp(beta0 + beta1*female*age + beta2*male*age);
   llike=log(eta*(count eq 0) + (1-eta)*pdf("poisson",count,mu));
   model dgeneral(llike);
run;
```

Posterior Estimates of the ZIP Model

The estimated $\pi(\beta_1 - \beta_2 > 0|\text{Data})$ is 100%, and on average, a female is expected to catch 1+ more fish than a male, with an increase of 10 years in age.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>bdif</td>
<td>10000</td>
<td>0.0165</td>
<td>0.00469</td>
<td>0.0133</td>
</tr>
<tr>
<td>mdif</td>
<td>10000</td>
<td>1.1811</td>
<td>0.0556</td>
<td>1.1423</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bdif</td>
<td>0.050</td>
<td>0.00739 0.0261</td>
<td>0.00641 0.0251</td>
</tr>
<tr>
<td>mdif</td>
<td>0.050</td>
<td>1.0767 1.2983</td>
<td>1.0662 1.2847</td>
</tr>
</tbody>
</table>
Posterior Marginal Density Plots

Posterior Density Plots

bdif

mdif
In each Markov chain iteration, PROC MCMC executes all programming statements once per observation in the input data set. Therefore, calculations that do not involve any data set variables are redundant and computationally inefficient.

Example:

```
bdif = beta1 - beta2;
mdif = exp(bdif * 10);
```
PROC MCMC *options*;

PARMS;

PARMS: } define parameters.

BEGINCNST;  
Programming statements;  
ENDCNST;

BEGINPRIOR;  
Programming statements;  
ENDPRIOR;

PRIOR: } declare prior distributions

Programming statements;  

MODEL  

Run;

set up statements, evaluated before but not during the simulation

calculate \( \log(\pi(\theta)) \) or \( g(\theta) \) (statements should not contain data set variables)

\( \log(f(y_i|\theta)) \), evaluated \( n \)-times per iteration
BEGINCNST/ENDCNST Statements

These statements jointly define a block, and the enclosed programming statements are processed only during the setup stage of a PROC MCMC run. These programming statements are not executed during the simulation and should be used to define constants or read data from a different data set. Example that defines a constant $\tau$:

```
begincnst;
  tau = 27;
endcnst;
```
BEGINPRIOR/ENDPRIOR Statements

These statements are designed to reduce unnecessary observation-level computations. They jointly define a block, and the enclosed programming statements are not executed for every data set observation. They are best reserved for calculations that relate to parameters only. Suppose that you parameterize your model on $\sigma$ but want to sample on the $\sigma^2$ scale:

```
parm s2;
beginprior;
  s = sqrt(s2);
endprior;
model y ~ normal(0, sd=s);
```

The calculation of $\sigma$ is identical for each observation.

BEGINPRIOR/ENDPRIOR ⇔ BEGINNODATA/ENDNODATA
BEGINPRIOR/ENDPRIOR Statements

You should not place data set variables within these segments. Suppose that \( y \) is a data set variable. The following calculation of \( \sigma \) will be incorrect because PROC MCMC does not know which value \( y \) should take while executing this statement.

```
beginprior;
  s = sqrt(y);
endprior;
```

You can include multiple PRIOR statements within the BEGINPRIOR/ENDPRIOR statements:

```
beginprior;
  prior alpha ~ n(0, var=10);
  prior beta ~ n(0, var=1);
endprior;
```

This does not change the way PROC MCMC understands your model.
The following program runs more efficiently:

```plaintext
proc mcmc data=catch seed=17 nmc=10000
   propcov=quanew plots=density monitor=(bdif mdif);
parms beta0 0 beta1 0 beta2 0;
parms eta .3;
prior beta: ~ normal(0,var=1000);
prior eta ~ uniform(0,1);
beginprior;
   bdif = beta1 - beta2;
   mdif = exp(bdif * 10);
endprior;
mu = exp(beta0 + beta1*female*age + beta2*male*age);
llike=log(eta*(count eq 0) + (1-eta)*pdf("poisson",count,mu));
model dgeneral(llike);
run;
```
Posterior Predictive Distribution

The posterior predictive distribution is the distribution of unobserved observations (prediction), conditional on the observed data. It is defined as:

\[
\pi(y_{\text{pred}}|y) = \int \pi(y_{\text{pred}}, \theta|y) d\theta
\]

\[
= \int \pi(y_{\text{pred}}|\theta, y)\pi(\theta|y) d\theta
\]

\[
= \int \pi(y_{\text{pred}}|\theta)\pi(\theta|y) d\theta
\]

The posterior predictive distribution can be seen as an integral of the likelihood function \( \pi(y_{\text{pred}}|\theta) \) with respect to the posterior \( \pi(\theta|y) \).
Posterior Predictive Distribution

You can use the posterior predictive distribution to:

- do prediction, using new covariates.
- check whether the model is consistent with the data. You generate samples from $\pi(y_{\text{pred}} | y)$ and see whether they differ systematically from the observed data (Gelman et al. 2004).

To generate a sample from the posterior predictive distribution, you draw a sample from the likelihood, conditional on the posterior samples of the parameters. In PROC MCMC, you do so by using the RAND function.
Binomial Model: Prediction

Suppose that in the binomial example, you want to know the following:

- predicted number of successes in a future trials of 242:

  \[ \text{event}_{\text{pred}} \sim \text{binomial}(242, p) \]

- probability that the number of successes exceeds a critical threshold:

  \[ \Pr \left( \text{event}_{\text{pred}} \geq 4 \right) \]
Binomial Model: Prediction

The following program fits a binomial model and makes prediction:

```plaintext
proc mcmc data=trials seed=17 nmc=10000
   monitor=(_parms_ xpred pcrit);
parm p;
prior p ~ beta(1,1);
beginprior;
   xpred = rand("binomial", p, 242);
   pcrit = (xpred > 4);
endprior;
model event ~ binomial(n,p);
run;
```

The `_parms_` symbol in the MONITOR= option is shorthand for all model parameters in the program.
## Prediction Based on the Binomial Model

### Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>p</td>
<td>10000</td>
<td>0.0181</td>
<td>0.00628</td>
<td>0.0135</td>
</tr>
<tr>
<td>xpred</td>
<td>10000</td>
<td>4.3062</td>
<td>2.4542</td>
<td>3.0000</td>
</tr>
<tr>
<td>pcrit</td>
<td>10000</td>
<td>0.4224</td>
<td>0.4940</td>
<td>0</td>
</tr>
</tbody>
</table>

### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.050</td>
<td>0.00795</td>
<td>0.0323</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00677</td>
<td>0.0305</td>
</tr>
<tr>
<td>xpred</td>
<td>0.050</td>
<td>0.00795</td>
<td>10.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00677</td>
<td>9.0000</td>
</tr>
<tr>
<td>pcrit</td>
<td>0.050</td>
<td>0.00795</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00677</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Outline

6 More examples

- Functions of parameters
- Posterior predictive distribution
- Incorporation of historical data (power prior)
- Sensitivity analysis
- Random-effects model
Incorporation of Historical Data

Suppose that, in addition to the data set trials in the binomial example, you have a data set from a pilot clinical trial:

```plaintext
data pilot;
  input event n;
datalines;
  5  163
;
```

noninformative analysis

```
prior p0 ~ beta(1,1);
model event ~ binomial(n,p0);
```

How would you capture the “information” contained in this trial on $p_0$ and incorporate it in the analysis using the current data set on $p$?
Power Prior

Power prior (Ibrahim and Chen 2000) enables you to retain information from a historical data set and use it as a prior distribution in the current analysis. To construct a power prior distribution, you can use the formula

\[ p(\theta|D_0, a_0) \propto L(\theta; D_0)^{a_0} \cdot \pi_0(\theta) \]

where

- \( D_0 = (n_0, y_0, X_0) \) is the historical (or pilot) data
- \( L(\theta; D_0) \) is the likelihood of \( \theta \) based on the historical data
- \( \pi_0(\theta) \) is the initial prior for \( \theta \), the prior for \( \theta \) before the historical data \( D_0 \) is observed
- \( a_0 \) is a discounting parameter constrained to \( 0 \leq a_0 \leq 1 \): \( a_0 = 0 \) corresponds to no incorporation of the historical data; \( a_0 = 1 \) corresponds to the Bayesian update of \( \pi(a_0) \).
Power Prior

The posterior distribution of $\theta$ is

$$p(\theta|D, D_0, a_0) \propto L(\theta; D) \cdot L(\theta; D_0)^{a_0} \cdot \pi_0(\theta)$$

$$\propto \prod_{i=1}^{n} f(y_i|\theta, x_i) \cdot \prod_{j=1}^{n_0} f(y_{0,j}|\theta, x_{0,j})^{a_0} \cdot \pi_0(\theta)$$

where

- $D = (n, y, X)$ are the data from the current study
- $y = \{y_i\}$ and $x = \{x_i\}$ for $i = 1 \cdots n$
- $y_{0} = \{y_{0,j}\}$ and $x_{0} = \{x_{0,j}\}$ for $j = 1 \cdots n_0$ are the historical data
- $f(\cdot)$ is the likelihood function for a single observation in either the historical or the current data
Power Prior

Combining the two data sets, you can form a new data set $D^*$ and rewrite the posterior distribution:

$$p(\theta|D^*, a_0) \propto \prod_{i=1}^{n+n_0} f_i(y_i|\theta, x_i) \cdot \pi_0(\theta)$$

where $f_i = \begin{cases} f(y_i|\theta, x_i) & \text{for each } i \text{ in the current data set} \\ f(y_{0,i}|\theta, x_{0,i})^{a_0} & \text{for each } i \text{ in the historical data set} \end{cases}$

You can use the IF statement to assign the appropriate likelihood function to different observations in the data set.
Binomial Model Using Power Prior

To fit a binomial model with power prior in PROC MCMC, you first want to combine both data sets and create a new group indicator variable:

```plaintext
data alldata;
   set trials(in=i) pilot;
   if i then group="current";
   else group="pilot";
run;
```

<table>
<thead>
<tr>
<th>OBS</th>
<th>event</th>
<th>n</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>86</td>
<td>current</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>69</td>
<td>current</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>71</td>
<td>current</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>113</td>
<td>current</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>103</td>
<td>current</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>163</td>
<td>pilot</td>
</tr>
</tbody>
</table>
Binomial Model: Power Prior

For each observation in the new data set alldata, the likelihood function is either a binomial (if `group == 'current'`) or a weighted binomial (if `group == 'pilot'`). The following SAS statements fit a power prior model:

```sas
proc mcmc data=alldata seed=17 nmc=50000 thin=5 outpost=aout;
  parm p 0.2;
  begincnst;
  a0 = 0.2;
  endcnst;
  prior p ~ beta(1,1);
  llike = logpdf("binomial", event, p, n);
  if (group = "pilot") then
    llike = a0 * llike;
  model general(llike);
run;
```
**Posterior Statistics of \( p \) Using Power Prior**

Posterior summary statistics reported here (with \( a_0 = 0.2 \)) are quite similar to analysis of the current data set without any information from the pilot study. Maybe a sensitivity analysis?

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>( p )</td>
<td>10000</td>
<td>0.0188</td>
<td>0.00625</td>
<td>0.0144</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0181</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0225</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>0.050</td>
<td>0.00876</td>
<td>0.0300</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00802</td>
<td>0.0317</td>
</tr>
</tbody>
</table>
Outline

More examples

- Functions of parameters
- Posterior predictive distribution
- Incorporation of historical data (power prior)
- Sensitivity analysis
- Random-effects model
Sensitivity analysis is the practice of understanding the variation and uncertainty of the posterior inferences as a result of a different prior or model (likelihood function) used in the analysis.

For example, you might want to compare power prior analyses of the binomial model with different weight values of $a_0$.

You can run PROC MCMC several times, each time with a different value of $a_0$, and compare the results. The BY statement makes it easy to carry out the analysis in one simple procedure call.
Sensitivity Analysis

Suppose that you want to compare three power prior models with values of \( a_0 \):

- \( a_0 = 0 \): an analysis that completely discards the pilot study
- \( a_0 = 0.5 \)
- \( a_0 = 1 \): an analysis that combines the pilot with the current study and assumes that they both come from the same population.
Sensitivity Analysis

The following statements generate a new data set with a BY-group indicator a0:

```sas
data sendata;
  set alldata;
  do a0 = 0, 0.5, 1;
    output;
  end;
proc sort;
  by a0;
run;
```
Sensitivity Analysis

The BY statement enables you to obtain separate analyses on data in groups defined by the BY variable. The following statements run three analyses, each with a different weight \( a_0 \):

```sas
proc mcmc data=sendata seed=17 nmc=10000 outpost=bout
diagnostics=none;
by a0;
parm p;
prior p ~ uniform(0, 1);
llike = logpdf("binomial", event, p, n);
if (group = 'pilot') then
   llike = a0 * llike;
model general(llike);
run;
```
Posterior Distributions of $p$ with Different Values of $a_0$
Outline

More examples

- Functions of parameters
- Posterior predictive distribution
- Incorporation of historical data (power prior)
- Sensitivity analysis
- Random-effects model
Random-Effects Model

Suppose that you no longer believe that all the trials share the same $p$ and you want to fit a random-effects model:

$$\text{event}_i \sim \text{binomial}(n_i, p_i)$$

where $i$ indexes the group. Two common choices for modeling $p_i$:

$$p_i \sim \text{beta}(a, b)$$
$$\text{logit}(p_i) \sim \text{normal}(\mu, \sigma^2)$$
Hyperparameters

If you choose constant values for $a$, $b$, $\mu$, or $\sigma^2$, you decide \textit{a priori} the amount of shrinkage you want on the $p_i$. For example:

- Choosing $a = 1$ and $b = 1$, or $\sigma^2 = \infty$, implies no shrinkage on the $p_i$. The random-effects model becomes an independent model (separate analysis).
- Choosing $\sigma^2 = 0$ imposes no variation amongst $p_i$. This reduces the random-effects model to the pooled model.

You can also use empirical Bayes estimates on the hyperparameters. This often gives posterior estimates that are similar to a full Bayes approach, if there are enough units or groups in the data to estimate the variance. But this plug-in method ignores uncertainty that your data indicates about the amount of shrinkage that should be used in the analysis.
Hyperprior Distributions

You can let the data decide what is the proper amount of shrinkage that should be used in estimating $p_i$—the right amount of strength you want to borrow from different groups to reduce variances. This amounts to placing hyperprior distributions on the hyperparameters.

The hyperprior distributions could potentially become very influential (to the posterior) in cases where the data contain little information to estimate the hyperparameters accurately. For example, see Spiegelhalter, Abrams, and Myles (2004), and Gelman et al. (2003) for discussions.

Strategies include:

- noninformative
- elicitation
- summary of evidence
Hyperprior Distributions

To illustrate the software, use proper but diffuse prior distributions on $a$ and $b$:

\[
\begin{align*}
\text{event}_i & \sim \text{binomial}(n_i, p_i) \\
p_i & \sim \text{beta}(a, b) \\
a, b & \sim \text{exponential}(\text{scale} = 100)
\end{align*}
\]
Binomial Model: Random-Effects Model

Here is the data again, and the pilot study is also included:

```plaintext
data trials;
  input event n center;
datalines;
2 86 1
2 69 2
1 71 3
1 113 4
1 103 5
5 163 6
;
```
Binomial Model: Random-Effects Model

The following SAS statements fit a hierarchical random-effects model, with a diffuse prior on the beta distribution:

```sas
proc mcmc data=trials nmc=100000 thin=10 outpost=outm;
   array p[6];
   parm p:;
   parm a b;
   prior a b ~ expon(scale=100);
   prior p: ~ beta(a, b);
   model event ~ binomial(n, p[center]);
run;
```

The ARRAY statement allocates an array of size 6 for the $p_i$s. In the MODEL statement, $p[\text{center}]$ is used to indicate the appropriate probability parameter for each of the observed responses.
95% HPD credible intervals and posterior point estimates for each of the six $p_i$. The solid line is the random-effects model; the dashed line is the independence model (individual analysis); the bottom line is the overall (pooled) estimates.
In Case You Are Interested:

If you wish to fit the following model:

\[ x_i \sim \text{binomial}(n_i, p_i) \]
\[ \gamma_i = \text{logit}(p_i) \sim \text{normal}(\mu, \sigma^2) \]
\[ \mu \sim \text{normal}(0, \text{precision} = 10^{-6}) \]
\[ \sigma^2 \sim \text{igamma}(0.001, \text{scale} = 0.001) \]

You can use the following statements in PROC MCMC:

```plaintext
array gamma[6];
parm gamma:
parm mu  s2;
prior mu ~ normal(0, prec=1e-6);
prior s2 ~ igamma(0.001, s=0.001);
prior gamma: ~ n(mu, sd=s2);
p = logistic(gamma[center]);
model event ~ binomial(n, p);
```
Case Control Study

Suppose that the binomial data come from the control group and the question of interest is on the treatment effect $\theta$. The complete data follow:

<table>
<thead>
<tr>
<th>OBS</th>
<th>ctrl</th>
<th>ctrlN</th>
<th>trt</th>
<th>trtN</th>
<th>center</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>86</td>
<td>2</td>
<td>70</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>69</td>
<td>3</td>
<td>94</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>71</td>
<td>0</td>
<td>78</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>113</td>
<td>2</td>
<td>127</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>103</td>
<td>1</td>
<td>91</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>163</td>
<td>7</td>
<td>142</td>
<td>6</td>
</tr>
</tbody>
</table>
Random-Effects Model: Odds Ratio

The random-effects model can be expressed as

\[\begin{align*}
\text{trt}_i &\sim \text{binomial}(\text{trtN}_i, p_i) \\
\text{ctrl}_i &\sim \text{binomial}(\text{ctrlN}_i, q_i) \\
\logit(p_i) &= \theta + \phi_i \\
\logit(q_i) &= \phi_i \\
\theta &\sim \text{normal}(0, \text{sd} = 10) \\
\phi_i &\sim \text{uniform}(-10, 10)
\end{align*}\]

where \( \phi_i \) is the baseline rate for the control group of trial \( i \) and \( \theta \) is the treatment effect or the log(odds ratio).
Random-Effects Model

The following SAS statements fit the desired random-effects model:

```sas
proc mcmc data=or monitor=( _parms_ OR );
   array phi[6];
   parm theta 0 phi: -5;
   prior theta ~ n(0, sd=10);
   prior phi: ~ uniform(-10, 10);
   p = logistic(theta + phi[center]);
   model trt ~ binomial(trtN, p);
   q = logistic(phi[center]);
   model ctrl ~ binomial(ctrlN, q);
   or = exp(theta);
run;
```

Multiple MODEL statements are allowed.
Posterior Estimates of Odds Ratio
Procedure Capability and Limitations

- The MCMC procedure carries the entire input data set (not the posterior draws) in memory. This largely dictates the maximum scope of the problem that the procedure can handle.

- The running time of a particular problem is approximately linear to the number of samples ($nsamples$), the number of simulations ($nsim$), and the number of parameters blocks in the program (number of PARMS statements, $nblocks$):

  \[ \text{run time} \propto nsamples \times nsim \times nblocks \]

- The faster a computer evaluates a single log-likelihood function, the faster the program runs:

  \[ \frac{200 \times 50,000 \times 3}{6 \times 10^6 \text{ per sec}} \approx 1 \text{ min} \]
Closing Remarks

PROC MCMC is a flexible procedure designed to fit a rich variety of linear and nonlinear, single-level and multilevel Bayesian models. The syntax is intuitive, and you can work with models that have standard or nonstandard distributions.

The flexibility of the procedure shifts certain responsibilities to the user. For example, you have to be careful about not constructing models that have improper posterior distributions.

Besides producing posterior estimates and convergence diagnostics, you can use PROC MCMC to compute functions of parameters, make predictions, compare models, and even incorporate user-defined sampling algorithms.

Bayesian computation is an active area of development, and we are interested in your feedback.
For More Information

See www.sas.com/statistics/ for:
- E-newsletter subscription
- News on updates and enhancements
- Examples library (Resources)

See www.support.sas.com/documentation/ for:
- Online SAS/STAT® documentation
- Downloadable chapter PDFs
References

Appendix
The Metropolis Algorithm

1. Let \( t = 0 \). Choose a starting point \( \theta^{(t)} \). This can be an arbitrary point as long as \( \pi(\theta^{(t)}|y) > 0 \).

2. Generate a new sample, \( \theta' \), from a proposal distribution \( q(\theta'|\theta^{(t)}) \).

3. Calculate the following quantity:

\[
r = \min \left\{ \frac{\pi(\theta'|y)}{\pi(\theta^{(t)}|y)}, 1 \right\}
\]

4. Sample \( u \) from the uniform distribution \( U(0, 1) \).

5. Set \( \theta^{(t+1)} = \theta' \) if \( u < r \); \( \theta^{(t+1)} = \theta^{(t)} \) otherwise.

6. Set \( t = t + 1 \). If \( t < T \), the number of desired samples, go back to Step 2; otherwise, stop.
The Metropolis Algorithm
The Metropolis Algorithm

\[ \pi(\theta|x) \]

\[ \theta' \sim N(\theta^{(0)}, \sigma) \]
The Metropolis Algorithm
The Metropolis Algorithm

If $\pi(\theta' | x) > \pi(\theta^{(0)} | x)$, $\theta^{(1)} = \theta'$

$\pi(\theta | x)$
The Metropolis Algorithm

If $\pi(\theta' | x) < \pi(\theta^{(0)} | x)$, accept

$\theta'$ with prob $\pi(\theta' | x) / \pi(\theta^{(0)} | x)$
The Metropolis Algorithm

\[ \theta' \sim N(\theta^{(1)}, \sigma) \]

\[ \pi(\theta|x) \]
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.
Bayesian Survival Analysis Procedures in SAS

Joseph G. Ibrahim

Department of Biostatistics

University of North Carolina
PROC LIFEREG provides Bayesian analysis methods for parametric survival models.

Distributions include

- Exponential
- Weibull
- Log-normal
- Normal
- 3-parameter gamma
- Logistic
- Log-logistic
The model parameters are the regression coefficients and a dispersion (precision, scale) parameter if the model has one.

The priors for the regression coefficients $\beta$ are allowed to be normal or uniform (improper).

The priors for the dispersion parameter $\sigma$ are either gamma, inverse gamma, or improper and of the form $p(\sigma) \propto \sigma^{-1}$.

The dispersion and regression parameters are assumed to be independent a priori.
LIFEREG formulates the survival models through an Accelerated Failure Time (AFT) approach in which the model is written as a linear model of the form

$$y = X\beta + \sigma\epsilon$$  \hspace{1cm} (1)

- The response vector, \(y\), can be untransformed implying \(y = T\), or transformed, leading to \(y = \log(T)\), where \(T\) is the survival time vector, which can be right censored, left censored, or interval censored.
- \(\beta\) is a \(p \times 1\) vector of regression coefficients.
- \(\sigma\) is a dispersion parameter.
LIFEREG

- $X$ is an $n \times p$ full-rank matrix of covariates.
- By default, SAS takes $y = \log(T)$.
- Different error distributions lead to the different models. For example, if we model $y = \log(T)$ via (1), and $\epsilon$ has an extreme value distribution, then the resulting model for $y$ is a Weibull model. This is the default.
- SAS allows you to pick the distribution for $y$ with a “LOG” or “NOLOG” option. The “LOG” option is the default.
### LIFEREG

<table>
<thead>
<tr>
<th>Distribution</th>
<th>NOLOG Specified?</th>
<th>Resulting Distribution for $y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPONENTIAL</td>
<td>No</td>
<td>Exponential</td>
</tr>
<tr>
<td>EXPONENTIAL</td>
<td>Yes</td>
<td>One-parameter extreme value</td>
</tr>
<tr>
<td>GAMMA</td>
<td>No</td>
<td>Generalized Gamma</td>
</tr>
<tr>
<td>GAMMA</td>
<td>Yes</td>
<td>Generalized gamma with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>untransformed responses</td>
</tr>
<tr>
<td>LOGISTIC</td>
<td>No</td>
<td>Logistic</td>
</tr>
<tr>
<td>LOGISTIC</td>
<td>Yes</td>
<td>Logistic (NOLOG has no effect)</td>
</tr>
<tr>
<td>LLOGISTIC</td>
<td>No</td>
<td>Log-logistic</td>
</tr>
<tr>
<td>LLOGISTIC</td>
<td>Yes</td>
<td>Logistic</td>
</tr>
<tr>
<td>LNORMAL</td>
<td>No</td>
<td>Lognormal</td>
</tr>
<tr>
<td>LNORMAL</td>
<td>Yes</td>
<td>Normal</td>
</tr>
<tr>
<td>NORMAL</td>
<td>No</td>
<td>Normal</td>
</tr>
<tr>
<td>NORMAL</td>
<td>Yes</td>
<td>Normal (NOLOG has no effect)</td>
</tr>
<tr>
<td>WEIBULL</td>
<td>No</td>
<td>Weibull</td>
</tr>
<tr>
<td>WEIBULL</td>
<td>Yes</td>
<td>Extreme value</td>
</tr>
</tbody>
</table>
The log-likelihood function of \((\beta, \sigma)\) for a set of right censored survival data is given by

\[
l = \log(L) = \sum_{i=1}^{n} \delta_i \log \left( \frac{f(u_i)}{\sigma} \right) + \sum_{i=1}^{n} (1 - \delta_i) \log(S(u_i)),
\]

where

\[
u_i = \frac{(y_i - \mathbf{x}_i' \beta)}{\sigma},
\]

and \(\delta_i = 1\) if the \(i\)th subject failed and \(\delta_i = 0\) if the \(i\)th subject was right censored.

\(f(u_i)\) is the density of \(\epsilon_i\), \(S(u_i)\) is the corresponding survival function, \(S(u_i) = 1 - F(u_i)\), where, \(F(u_i)\) is the cdf of \(\epsilon_i\).
The BAYES Statement

- Bayesian analyses in GENMOD, LIFEREG, and PHREG are all facilitated through the BAYES statement.
- These are shared options across all of the procedures:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INITIAL =</td>
<td>initial values of the chain</td>
</tr>
<tr>
<td>NBI =</td>
<td>number of burn-in iterations</td>
</tr>
<tr>
<td>NMC =</td>
<td>number of iterations after burn-in</td>
</tr>
<tr>
<td>SEED =</td>
<td>random number generator seed</td>
</tr>
<tr>
<td>THINNING =</td>
<td>thinning of the Markov chain</td>
</tr>
<tr>
<td>DIAGNOSTICS =</td>
<td>convergence diagnostics</td>
</tr>
<tr>
<td>PLOTS =</td>
<td>diagnostic plots</td>
</tr>
<tr>
<td>SUMMARY =</td>
<td>summary statistics</td>
</tr>
<tr>
<td>COEFFPRIOR =</td>
<td>prior for the regression coefficients</td>
</tr>
</tbody>
</table>
The `LIFEREG` specific options are:

<table>
<thead>
<tr>
<th>EXPONENTIALSCALEPRIOR =</th>
<th>prior for the exponential scale parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCALEPRIOR =</td>
<td>prior for the scale parameter</td>
</tr>
<tr>
<td>WEIBULLSCALEPRIOR =</td>
<td>prior for the Weibull scale parameter</td>
</tr>
<tr>
<td>WEIBULLSHAPEPRIOR =</td>
<td>prior for the Weibull shape parameter</td>
</tr>
</tbody>
</table>

The Weibull shape parameter is $\gamma \equiv \sigma^{-1}$. 
For the Weibull regression model, the likelihood function of \((\beta, \sigma^{-1})\) is constructed by letting \(\mu = \mathbf{x}'\beta\) according to the LIFEREG manual.

The Weibull survival density for \(T_i\) is given by

\[
f(t_i) = \sigma^{-1} \exp(-\mathbf{x}_i'\beta/\sigma) \ t_i^{1/\sigma-1} \exp\left\{ -t_i^{1/\sigma} \exp(-\mathbf{x}_i'\beta/\sigma) \right\}.
\]

The Weibull survival function is given by

\[
S(t_i) = \exp\left\{ -\exp(-\mathbf{x}_i'\beta/\sigma) \ t_i^{1/\sigma} \right\}.
\]
The log-likelihood for the Weibull regression model with right censored data is thus given by

\[ l(\beta, \sigma^{-1}) = \sum_{i=1}^{n} \delta_i \log(f(t_i)) + \sum_{i=1}^{n} (1 - \delta_i) \log(S(t_i)), \quad (2) \]

where \( \delta_i = 1 \) if the \( i \)th subject failed and \( \delta_i = 0 \) if the \( i \)th subject was right censored.

The joint posterior distribution of \((\beta, \sigma^{-1})\) is thus given by

\[ p(\beta, \sigma^{-1}|\text{Data}) \propto L(\beta, \sigma^{-1})\pi(\beta)\pi(\sigma^{-1}), \quad (3) \]

where \( \pi(\beta) \) can be chosen as normal or uniform and \( \pi(\sigma^{-1}) \) can be chosen as gamma or improper.
The exponential model is a special case of the Weibull model obtained by setting $\sigma = 1$.

The lognormal survival density for $T_i$ is given by

$$f(t_i) = \frac{1}{\sqrt{2\pi}\sigma t_i} \exp \left( -\frac{1}{2} \left( \frac{\log(t_i) - x_i'\beta}{\sigma} \right)^2 \right).$$

The lognormal survival function is given by

$$S(t_i) = 1 - \Phi \left( \frac{\log(t_i) - x_i'\beta}{\sigma} \right).$$
The log-likelihood for the lognormal regression model with right censored data is thus given by

\[
l(\beta, \sigma^{-1}) = \sum_{i=1}^{n} \delta_i \log(f(t_i)) + \sum_{i=1}^{n} (1 - \delta_i) \log(S(t_i)), \quad (4)\]

where \(\delta_i = 1\) if the \(ith\) subject failed and \(\delta_i = 0\) if the \(ith\) subject was right censored.
We consider the E1684 melanoma clinical trial with $n = 286$ subjects.

We consider fitting a parametric survival model with three covariates: Treatment (trt) (High-dose IFN and Observation), age, and sex.

We consider the Weibull model, exponential model, and lognormal model.
data melanoma;
    infile 'H:/e1684ws.dat';
    input case study age trt sex perform nodes breslow failtime rfscens survtime scens;
run;

proc lifereg;
    model failtime*rfscens(0) = trt age sex /dist=weibull;
    bayes seed = 5432 nbi=3000 nmc = 20000
    WeibullShapePrior=gamma
    coeffprior = uniform
    diagnostics = all plots=all;
run;

proc lifereg;
    model failtime*rfscens(0) = trt age sex /dist=exponential;
    bayes seed = 5432 nbi=3000 nmc = 20000
    coeffprior = uniform
    diagnostics = all plots=all;
run;
SAS Code for LIFEREG: E1684 Data

```sas
proc lifereg;
  model failtime*rfscens(0) = trt age sex /dist=weibull;
  bayes seed = 5432 nbi=3000 nmc = 20000
  WeibullShapePrior=gamma(shape=0.01,iscale=.01)
  coeffprior = normal
  diagnostics = all plots = all;

proc lifereg;
  model failtime*rfscens(0) = trt age sex /dist=Lognormal;
  bayes seed = 5432 nbi=3000 nmc = 20000
  coeffprior = uniform
  diagnostics = all plots = all;

run;
```
Bayesian Analysis

Model Information

Data Set                  WORK.MELANOMA
Dependent Variable       Log(failtime)
Censoring Variable       rfscens
Censoring Value(s)       0
Number of Observations   284
Noncensored Values       196
Right Censored Values    88
Left Censored Values     0
Interval Censored Values 0
Zero or Negative Response 1
Burn-In Size             3000
MC Sample Size           20000
Thinning                 1
Name of Distribution     Weibull
Log Likelihood           -516.1569613

Number of Observations Read  286
Number of Observations Used  284
Missing Values              1
Weibull Model for E1684 Data: Default Priors

Algorithm converged.

Analysis of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.5883</td>
<td>0.4778</td>
<td>0.6518</td>
</tr>
<tr>
<td>trt</td>
<td>1</td>
<td>0.6400</td>
<td>0.2459</td>
<td>0.1582</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>-0.0115</td>
<td>0.0090</td>
<td>-0.0292</td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>0.0011</td>
<td>0.2503</td>
<td>-0.4896</td>
</tr>
<tr>
<td>Scale</td>
<td>1</td>
<td>1.7053</td>
<td>0.1025</td>
<td>1.5158</td>
</tr>
<tr>
<td>Weibull Shape</td>
<td>1</td>
<td>0.5864</td>
<td>0.0352</td>
<td>0.5213</td>
</tr>
</tbody>
</table>
Weibull Model for E1684 Data: Default Priors

The LIFEREG Procedure

Bayesian Analysis

Uniform Prior for Regression Coefficients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Constant</td>
</tr>
<tr>
<td>trt</td>
<td>Constant</td>
</tr>
<tr>
<td>age</td>
<td>Constant</td>
</tr>
<tr>
<td>sex</td>
<td>Constant</td>
</tr>
</tbody>
</table>

Independent Prior Distributions for Model Parameters

<table>
<thead>
<tr>
<th>Prior Parameter</th>
<th>Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull Shape</td>
<td>Gamma</td>
<td>Shape 0.001 Inverse Scale 0.001</td>
</tr>
</tbody>
</table>

Initial Values of the Chain

<table>
<thead>
<tr>
<th>Chain</th>
<th>Seed</th>
<th>Intercept</th>
<th>trt</th>
<th>age</th>
<th>sex</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5432</td>
<td>1.589109</td>
<td>0.641521</td>
<td>-0.01153</td>
<td>0.001324</td>
<td>0.584295</td>
</tr>
<tr>
<td>2</td>
<td>3.261013</td>
<td>-0.09873</td>
<td>-0.03875</td>
<td>-0.75242</td>
<td>0.48771</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.08279</td>
<td>1.381771</td>
<td>0.015685</td>
<td>0.755069</td>
<td>0.700007</td>
<td></td>
</tr>
</tbody>
</table>

Fit Statistics

- AIC (smaller is better) 1042.314
- AICC (smaller is better) 1042.530
- BIC (smaller is better) 1060.559
- DIC (smaller is better) 1042.377
- pD (effective number of parameters) 4.988
Weibull Model for E1684 Data: Default Priors

The LIFEREG Procedure

Bayesian Analysis

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20000</td>
<td>1.6103</td>
<td>0.4878</td>
<td>1.2806</td>
<td>1.6075</td>
<td>1.9352</td>
</tr>
<tr>
<td>trt</td>
<td>20000</td>
<td>0.6491</td>
<td>0.2511</td>
<td>0.4806</td>
<td>0.6472</td>
<td>0.8168</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>-0.0116</td>
<td>0.00928</td>
<td>-0.0179</td>
<td>-0.0117</td>
<td>-0.00539</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>0.00645</td>
<td>0.2562</td>
<td>-0.1672</td>
<td>0.00304</td>
<td>0.1771</td>
</tr>
<tr>
<td>WeibShape</td>
<td>20000</td>
<td>0.5777</td>
<td>0.0349</td>
<td>0.5539</td>
<td>0.5772</td>
<td>0.6008</td>
</tr>
</tbody>
</table>

Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>0.6701 2.5729</td>
<td>0.6775 2.5769</td>
</tr>
<tr>
<td>trt</td>
<td>0.050</td>
<td>0.1622 1.1414</td>
<td>0.1513 1.1281</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.0298 0.00667</td>
<td>-0.0300 0.00631</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.4879 0.5205</td>
<td>-0.4890 0.5180</td>
</tr>
<tr>
<td>WeibShape</td>
<td>0.050</td>
<td>0.5114 0.6474</td>
<td>0.5115 0.6474</td>
</tr>
</tbody>
</table>
Weibull Model for E1684 Data: Default Priors

The LIFEREG Procedure

Bayesian Analysis

Posterior Autocorrelations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lag 1</th>
<th>Lag 5</th>
<th>Lag 10</th>
<th>Lag 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.0034</td>
<td>-0.0028</td>
<td>0.0004</td>
<td>-0.0006</td>
</tr>
<tr>
<td>trt</td>
<td>0.0265</td>
<td>-0.0020</td>
<td>-0.0054</td>
<td>0.0127</td>
</tr>
<tr>
<td>age</td>
<td>0.0074</td>
<td>-0.0043</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
<tr>
<td>sex</td>
<td>0.0607</td>
<td>-0.0019</td>
<td>0.0012</td>
<td>-0.0058</td>
</tr>
<tr>
<td>WeibShape</td>
<td>0.0305</td>
<td>0.0064</td>
<td>-0.0059</td>
<td>-0.0071</td>
</tr>
</tbody>
</table>
Weibull Model for E1684 Data: Default Priors

Gelman-Rubin Diagnostics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.0000</td>
<td>1.0001</td>
</tr>
<tr>
<td>trt</td>
<td>1.0000</td>
<td>1.0001</td>
</tr>
<tr>
<td>age</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>sex</td>
<td>1.0000</td>
<td>1.0002</td>
</tr>
<tr>
<td>WeibShape</td>
<td>1.0001</td>
<td>1.0003</td>
</tr>
</tbody>
</table>

Geweke Diagnostics

| Parameter   | z        | Pr > |z|   |
|-------------|----------|------|----|
| Intercept   | 0.1375   | 0.8906 |
| trt         | -0.9739  | 0.3301 |
| age         | 0.1603   | 0.8726 |
| sex         | -0.1545  | 0.8772 |
| WeibShape   | -0.3970  | 0.6914 |
Weibull Model for E1684 Data: Default Priors

Raftery-Lewis Diagnostics
Quantile=0.025 Accuracy=+-0.005 Probability=0.95 Epsilon=0.001

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Burn-in</th>
<th>Total</th>
<th>Minimum</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.</td>
<td>.</td>
<td>3746</td>
<td></td>
</tr>
<tr>
<td>trt</td>
<td>.</td>
<td>.</td>
<td>3746</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>2</td>
<td>3818</td>
<td>3746</td>
<td>1.0192</td>
</tr>
<tr>
<td>sex</td>
<td>2</td>
<td>3850</td>
<td>3746</td>
<td>1.0278</td>
</tr>
<tr>
<td>WeibShape</td>
<td>2</td>
<td>3749</td>
<td>3746</td>
<td>1.0008</td>
</tr>
</tbody>
</table>
The LIFEREG Procedure

Bayesian Analysis

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interception</td>
<td>20000</td>
<td>1.5979</td>
<td>0.2766</td>
<td>1.4094</td>
<td>1.5962</td>
<td>1.7842</td>
</tr>
<tr>
<td>trt</td>
<td>20000</td>
<td>0.4509</td>
<td>0.1440</td>
<td>0.3542</td>
<td>0.4503</td>
<td>0.5471</td>
</tr>
<tr>
<td>Age</td>
<td>20000</td>
<td>-0.00916</td>
<td>0.00523</td>
<td>-0.0127</td>
<td>-0.00914</td>
<td>-0.00562</td>
</tr>
<tr>
<td>Sex</td>
<td>20000</td>
<td>-0.0249</td>
<td>0.1479</td>
<td>-0.1248</td>
<td>-0.0260</td>
<td>0.0742</td>
</tr>
</tbody>
</table>

Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interception</td>
<td>0.050</td>
<td>1.0589 2.1387</td>
<td>1.0675 2.1433</td>
</tr>
<tr>
<td>trt</td>
<td>0.050</td>
<td>0.1690 0.7314</td>
<td>0.1685 0.7305</td>
</tr>
<tr>
<td>Age</td>
<td>0.050</td>
<td>-0.0194 0.00101</td>
<td>-0.0196 0.000742</td>
</tr>
<tr>
<td>Sex</td>
<td>0.050</td>
<td>-0.3112 0.2677</td>
<td>-0.3085 0.2700</td>
</tr>
</tbody>
</table>
Weibull Model with Normal and Gamma Priors

The LIFEREG Procedure

Bayesian Analysis

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20000</td>
<td>1.6103</td>
<td>0.4878</td>
<td>1.2806</td>
<td>1.6075</td>
<td>1.9351</td>
</tr>
<tr>
<td>trt</td>
<td>20000</td>
<td>0.6491</td>
<td>0.2511</td>
<td>0.4806</td>
<td>0.6472</td>
<td>0.8168</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>-0.0116</td>
<td>0.00928</td>
<td>-0.0179</td>
<td>-0.0117</td>
<td>-0.00539</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>0.00645</td>
<td>0.2562</td>
<td>-0.1672</td>
<td>0.00304</td>
<td>0.1771</td>
</tr>
<tr>
<td>WeibShape</td>
<td>20000</td>
<td>0.5777</td>
<td>0.0349</td>
<td>0.5539</td>
<td>0.5772</td>
<td>0.6008</td>
</tr>
</tbody>
</table>
## Weibull Model with Normal and Gamma Priors

### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>0.6701</td>
<td>2.5729</td>
</tr>
<tr>
<td>trt</td>
<td>0.050</td>
<td>0.1622</td>
<td>1.1413</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.0298</td>
<td>0.00667</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.4879</td>
<td>0.5205</td>
</tr>
<tr>
<td>WeibShape</td>
<td>0.050</td>
<td>0.5114</td>
<td>0.6474</td>
</tr>
</tbody>
</table>
E1684 Data: Lognormal Model with Uniform Prior

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>0.6701</td>
<td>2.5729</td>
</tr>
<tr>
<td>trt</td>
<td>0.050</td>
<td>0.1622</td>
<td>1.1413</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.0298</td>
<td>0.00667</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.4879</td>
<td>0.5205</td>
</tr>
<tr>
<td>WeibShape</td>
<td>0.050</td>
<td>0.5114</td>
<td>0.6474</td>
</tr>
</tbody>
</table>
### E1684 Data: Lognormal Model with Uniform Prior

The LIFEREG Procedure

Bayesian Analysis

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20000</td>
<td>0.6577</td>
<td>0.4992</td>
<td>0.3212</td>
<td>0.6536</td>
<td>0.9951</td>
</tr>
<tr>
<td>trt</td>
<td>20000</td>
<td>0.6710</td>
<td>0.2463</td>
<td>0.5079</td>
<td>0.6703</td>
<td>0.8339</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>-0.00895</td>
<td>0.00960</td>
<td>-0.0153</td>
<td>-0.00898</td>
<td>-0.00251</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>0.00341</td>
<td>0.2549</td>
<td>-0.1675</td>
<td>0.00290</td>
<td>0.1750</td>
</tr>
<tr>
<td>Scale</td>
<td>20000</td>
<td>1.9971</td>
<td>0.1112</td>
<td>1.9194</td>
<td>1.9911</td>
<td>2.0687</td>
</tr>
</tbody>
</table>
## E1684 Data: Lognormal Model with Uniform Prior

### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.050</td>
<td>-0.3179</td>
<td>1.6451</td>
</tr>
<tr>
<td>trt</td>
<td>0.050</td>
<td>0.1952</td>
<td>1.1613</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.0278</td>
<td>0.00989</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.4953</td>
<td>0.5062</td>
</tr>
<tr>
<td>Scale</td>
<td>0.050</td>
<td>1.7942</td>
<td>2.2287</td>
</tr>
</tbody>
</table>
There are many other options in LIFEREG:

```sas
proc lifereg;
  model failtime*rfscens(0) = trt age sex /dist=exponential;
  bayes seed = 5432 nbi=3000 nmc = 20000
  coeffprior = uniform
  diagnostics = all
  summaries = corr
  thinning =10 plots=all;
```

- **WeibullShapePrior=gamma(relshape=c)** results in a gamma(\(c\hat{\beta}, c\)) prior for \(\sigma^{-1}\) in the Weibull model.
- **WeibullShapePrior=gamma(shape=c)** results in a gamma(\(c, c\)) prior.
data NormalPrior1;
input _type_ $ Intercept trt age sex;
datalines;
  Var 100 100 100 100
  Mean 1.0 2.0 3.0 4.0
;
run;

proc lifereg data=melanoma;
  model failtime*rfscens(0) = trt age sex /dist=exponential;
  bayes seed = 5432 nbi=3000 nmc = 20000
  coeffprior = normal(input=NormalPrior1);
run;
data NormalPrior2;
input _type_ $ _name_:$9. Intercept trt age sex;
cards;
Mean . 1 2 3 4
Cov Intercept 2 1 1 1
Cov trt 1 3 1 1
Cov age 1 1 4 1
Cov sex 1 1 1 5
;
run;
PHREG fits semiparametric survival models using

- Cox regression models based on partial likelihood: exact/Breslow/Efron/discrete logistic likelihood, time independent, time dependent, with/without ties in time. Thus, the partial likelihood itself is treated as the likelihood.
- Piecewise exponential models: on both the baseline hazard and logarithm of the baseline hazard.

It does not fit models with certain data constraints, as for example, data with recurrent events.
Bayesian Inference Based on Partial Likelihood

- The Breslow partial likelihood is given by
  \[
  L_P(\beta) = \prod_{i=1}^{k} \frac{\exp\left(\beta' \sum_{j \in D_d} x_j(t_i)\right)}{\left[\sum_{l \in R_i} \exp(\beta' x_l(t_i))\right]^{d_i}},
  \]
  where \(x_l(t)\) is the vector of covariates for the \(lth\) individual at time \(t\), \(t_1 < t_2 < \ldots < t_k\) denote the \(k\) distinct failure (event) times, and \(d_i\) denotes the multiplicity of failure times at \(t_i\).

- Let the data be denoted by \(D = \{(t_i, x_i, \delta_i), i = 1, \ldots, n\}\). The posterior distribution of \(\beta\) based on \(L_P(\beta)\) is given by
  \[
  p(\beta|D) \propto L_P(\beta)\pi(\beta).
  \]

- For the partial likelihood model, \(\pi(\beta)\) is allowed to be multivariate normal or uniform.
We first partition the time axis. Let

\[ a_0 = 0 < a_1 < a_2 < \ldots a_{J-1} < a_J = \infty \]

be a partition of the time axis.

The hazard for subject \( i \) is

\[ h(t|x_i, \beta) = h_0(t) \exp(x_i^\prime \beta), \]

where

\[ h_0(t) = \lambda_j, \quad a_{j-1} < t < a_j \quad (j = 1, \ldots, J). \]

Thus, the hazard function for subject \( i \) in the \( jth \) time interval is

\[ h(t|x_i) = \lambda_j \exp(x_i^\prime \beta), \quad a_{j-1} < t < a_j. \] (5)
From the hazard function, we can define the baseline cumulative hazard function, which is given by

\[ H_0(t) = \sum_{j=1}^{J} \lambda_j \Delta_j(t), \]

where

\[ \Delta_j(t) = \begin{cases} 
0 & t < a_{j-1} \\
 t - a_{j-1} & a_{j-1} \leq t < a_j \\
a_j - a_{j-1} & t \geq a_j
\end{cases}. \]
Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$. The log-likelihood of $(\boldsymbol{\lambda}, \boldsymbol{\beta})$ is given by

$$l(\beta, \lambda) = \sum_{i=1}^{n} \delta_i \left[ \sum_{j=1}^{J} I(a_{j-1} \leq t_i < a_j) \log(\lambda_j) + \mathbf{x}_i' \beta \right] - \sum_{i=1}^{n} \left[ \sum_{j=1}^{J} \Delta_j(t_i) \lambda_j \right] \exp(\mathbf{x}_i' \beta)$$

$$= \sum_{j=1}^{J} d_j \log(\lambda_j) + \sum_{i=1}^{n} \delta_i \mathbf{x}_i' \beta - \sum_{j=1}^{J} \lambda_j \left[ \sum_{i=1}^{n} \Delta_j(t_i) \exp(\mathbf{x}_i' \beta) \right],$$

where $d_j = \sum_{i=1}^{n} \delta_i I(a_{j-1} \leq t_i < a_j)$. 

### Piecewise Constant Baseline Hazard Model
We note that for $1 \leq j \leq J$, the full conditional distribution for $\lambda_j$ is log-concave only when $d_j > 0$, but the full conditionals of the $\beta_k$'s are always log-concave.

The MLE of $\beta$ can be obtained via Newton-Raphson methods. The MLE of $\lambda$ has a closed form depending on $\beta$.

Specifically, for a given $\beta$, $\frac{\partial l}{\partial \lambda} = 0$ gives

$$
\tilde{\lambda}_j(\beta) = \frac{d_j}{\sum_{i=1}^{n} \Delta_j(t_i) \exp(x_i'\beta)}, \quad j = 1, \ldots, J.
$$
Substituting these values into $l(\lambda, \beta)$ gives the profile log-likelihood of $\beta$:

$$l_p(\beta) = \sum_{i=1}^{n} \delta_i x_i' \beta - \sum_{j=1}^{J} d_j \log \left[ \sum_{l=1}^{n} \Delta_j(t_l) \exp(x_l' \beta) \right] + c,$$  \hspace{1cm} (6)

where $c = \sum_{j=1}^{J} (d_j \log(d_j) - d_j)$.

Since the constant $c$ does not depend on $\beta$, it can be discarded from $l_p(\beta)$ in the optimization.

The MLE of $\beta$ is thus found by maximizing

$$l_p(\beta) = \sum_{i=1}^{n} \delta_i x_i' \beta - \sum_{j=1}^{J} d_j \log \left[ \sum_{l=1}^{n} \Delta_j(t_l) \exp(x_l' \beta) \right]$$

with respect to $\beta$. 

---

**Piecewise Constant Baseline Hazard Model**
The MLE of $\lambda$ is given by

$$\hat{\lambda} = \tilde{\lambda}(\hat{\beta}).$$

Let

$$S_j^{(r)}(\beta) = \sum_{l=1}^{n} \Delta_j(t_l) \exp(x_l'\beta)x_l^{\otimes r},$$

for $r = 0, 1, 2$, and $j = 1, \ldots, J$.

$x_l^{\otimes 0} = 1$, $x_l^{\otimes 1} = x_l$, and $x_l^{\otimes 2} = x_l x_l'$.

Also, let

$$E_j(\beta) = \frac{S_j^{(1)}(\beta)}{S_j^{(0)}(\beta)}.$$
The partial derivatives of $l_p(\beta)$ are

$$\frac{\partial l_p(\beta)}{\partial \beta} = \sum_{i=1}^{n} \delta_i x_i - \sum_{j=1}^{J} d_j E_j(\beta)$$

$$\frac{\partial^2 l_p(\beta)}{\partial \beta^2} = \sum_{j=1}^{J} d_j \left\{ \frac{S_j^{(2)}(\beta)}{S_j^{(0)}(\beta)} - [E_j(\beta)] [E_j(\beta)]' \right\}$$
The asymptotic covariance matrix for \((\hat{\lambda}, \hat{\beta})\) is obtained as the inverse of the information matrix given by

\[
- \frac{\partial^2 l(\hat{\lambda}, \hat{\beta})}{\partial \lambda^2} = \text{diag} \left( \frac{d_1}{\hat{\lambda}_1^2}, \ldots, \frac{d_J}{\hat{\lambda}_J^2} \right)
\]

\[
- \frac{\partial^2 l(\hat{\lambda}, \hat{\beta})}{\partial \beta^2} = \sum_{j=1}^{J} \hat{\lambda}_j S_j^{(2)}(\hat{\beta})
\]

\[
- \frac{\partial^2 l(\hat{\lambda}, \hat{\beta})}{\partial \lambda \beta} = (S_1^{(1)}(\hat{\beta}), \ldots, S_J^{(1)}(\hat{\beta})).
\]
Piecewise Constant Baseline Hazard Model

- By letting $\alpha_j = \log(\lambda_j)$, $j = 1, \ldots, J$, we can build prior correlation between the $\lambda_j$'s by specifying a multivariate normal prior for $\alpha = (\alpha_1, \ldots, \alpha_J)$, that is,

  $$\alpha \sim N(\alpha_0, \Sigma_\alpha).$$

- In this case, the log-likelihood of $(\alpha, \beta)$ is given by

  $$l(\alpha, \beta) = \sum_{j=1}^{J} d_j \alpha_j + \sum_{i=1}^{n} \delta_i x'_i \beta - \sum_{j=1}^{J} \exp(\alpha_j) S_j^{(0)}(\beta).$$

- The MLE of $\lambda_j$ is given by

  $$\exp(\hat{\alpha}_j) = \hat{\lambda}_j = \frac{d_j}{S_j^{(0)}(\hat{\beta})}.$$
We note that the full conditionals for the $\alpha_j$’s and $\beta_k$’s are always log-concave.

The asymptotic covariance matrix of $(\hat{\alpha}, \hat{\beta})$ is obtained as the inverse of the information matrix formed by

$$
- \frac{\partial^2 l(\hat{\alpha}, \hat{\beta})}{\partial \alpha^2} = \operatorname{diag} \left( \exp(\hat{\alpha}_1) S_1^{(0)}(\hat{\beta}), \ldots, \exp(\hat{\alpha}_J) S_J^{(0)}(\hat{\beta}) \right)
$$

$$
- \frac{\partial^2 l(\hat{\alpha}, \hat{\beta})}{\partial \beta^2} = \sum_{j=1}^{J} \exp(\hat{\alpha}_j) S_j^{(2)}(\hat{\beta})
$$

$$
- \frac{\partial^2 l(\hat{\alpha}, \hat{\beta})}{\partial \alpha \beta} = \left( \exp(\hat{\alpha}_1) S_1^{(1)}(\hat{\beta}), \ldots, \exp(\hat{\alpha}_J) S_J^{(1)}(\hat{\beta}) \right).
$$
Summary of PHREG

- Partial Likelihood Model (Cox): Uniform improper prior and normal priors on the regression coefficients are allowed.

- Piecewise Exponential:
  - Regression coefficients $\beta$: uniform improper and normal priors.
  - log hazard parameters $\alpha = \log(\lambda)$: uniform improper and normal priors.
  - Regression parameters and log hazard parameters: $(\beta, \alpha)$ can be taken jointly as multivariate normal and need not be independent a priori.
  - hazard parameters $\lambda$: Jeffreys’s (improper), uniform improper, product of independent gamma densities, and AR(1) gamma priors.
Priors for PHREG

- The Jeffreys’s prior for $\lambda$ is

$$
\pi(\lambda) \propto \prod_{j=1}^{J} \frac{1}{\lambda_j}.
$$

- The uniform improper prior for $\lambda$ is

$$
\pi(\lambda) \propto 1.
$$

- For the independent gamma priors, we assume $\lambda_j \sim \text{gamma}(a_j, b_j)$, and independent for $j = 1, \ldots, J$. The prior for $\lambda$ is then given by

$$
\pi(\lambda) \propto \prod_{j=1}^{J} \left\{ \lambda_j^{a_j-1} \exp(-\lambda_j b_j) \right\}.
$$
Priors for PHREG

- The AR(1) gamma prior for $\lambda$ is given by

$$
\begin{align*}
\lambda_1 & \sim \text{gamma}(a_1, b_1) \\
\lambda_2 | \lambda_1 & \sim \text{gamma} \left( a_2, \frac{b_2}{\lambda_1} \right) \\
& \ldots \\
\lambda_J | \lambda_{J-1} & \sim \text{gamma} \left( a_J, \frac{b_J}{\lambda_{J-1}} \right).
\end{align*}
$$

- The prior density for $\lambda$ is then given by

$$
\pi(\lambda) \propto \lambda_1^{a_1-1} \exp(-b_1 \lambda_1) \prod_{j=2}^{J} \left( \frac{b_j}{\lambda_{j-1}} \right)^{a_j} \lambda_j^{a_j-1} \exp \left( -\frac{b_j \lambda_j}{\lambda_{j-1}} \right).
$$
Priors for PHREG

* For $\alpha$, we can take
  \[ \pi(\alpha) \propto 1, \]
  \[ \pi(\alpha) \propto \exp \left[ -\frac{1}{2} (\alpha - \alpha_0)' \Phi_0^{-1} (\alpha - \alpha_0) \right]. \]

* We can also take a joint multivariate normal prior for $(\alpha, \beta)$, given by
  \[ \pi(\alpha, \beta) \propto \exp \left( -\frac{1}{2} \left[ (\alpha - \alpha_0)', (\beta - \beta_0)' \right] \Phi_0^{-1} \left[ (\alpha - \alpha_0)', (\beta - \beta_0)' \right] \right). \]
Priors for PHREG

For the piecewise exponential model, the joint posterior of $(\lambda, \beta)$ is now given by

$$ p(\lambda, \beta|D) \propto L(\lambda, \beta)\pi(\beta)\pi(\lambda). $$

For the loghazard model, the joint posterior of $(\alpha, \beta)$ is given by

$$ p(\alpha, \beta|D) \propto \begin{cases} L(\alpha, \beta)\pi(\alpha, \beta) & \text{if } (\alpha, \beta) \sim MVN \\ L(\alpha, \beta)\pi(\alpha)\pi(\beta) & \text{otherwise.} \end{cases} $$
The key command for the piecewise exponential model is

\[
\text{piecewise} = \text{details of the piecewise exponential model}
\]

The user can specify the number of intervals (hazard parameters \( \lambda_j \)), or interval partitions, and the prior distribution for the hazard functions.

The default number of intervals is \( J = 8 \).
PHREG for E1684 Data

data melanoma;
  infile 'H:/e1684ws.dat';
  input case study age trt sex perform nodes breslow failtime rfscens
    survtime scens;
  run;

proc phreg;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 532 nbi=3000 nmc = 20000
    coeffprior = uniform
    diagnostics = none;
  run;

proc phreg;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
    coeffprior = normal
    diagnostics = none;
  run;
PHREG for E1684 Data

proc phreg;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
    coeffprior = normal
    piecewise = hazard(N=15 prior=gamma)
  diagnostics = none;
run;

proc phreg;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
    coeffprior = normal
    piecewise = hazard(N=15 prior=improper)
  diagnostics = none;
run;
proc phreg;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
    coeffprior = normal
    piecewise = loghazard(N=15 prior=normal)
  diagnostics = none;
run;
Cox Model for E1684 Data: Uniform Priors

The PHREG Procedure

Bayesian Analysis

Model Information

Data Set WORK.MELANOMA
Dependent Variable failtime
Censoring Variable rfscens
Censoring Value(s) 0
Model Cox
Ties Handling BRESLOW
Burn-In Size 3000
MC Sample Size 20000
Thinning 1

Number of Observations Read 286
Number of Observations Used 285
Cox Model for E1684 Data: Uniform Priors

Summary of the Number of Event and Censored Values

<table>
<thead>
<tr>
<th>Total</th>
<th>Event</th>
<th>Censored</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>285</td>
<td>196</td>
<td>89</td>
<td>31.23</td>
</tr>
</tbody>
</table>

Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>1</td>
<td>-0.3598</td>
<td>0.1437</td>
<td>-0.6415 -0.0782</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>0.00491</td>
<td>0.00532</td>
<td>-0.00551 0.0153</td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>-0.0180</td>
<td>0.1469</td>
<td>-0.3059 0.2698</td>
</tr>
</tbody>
</table>

Uniform Prior for Regression Coefficients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>Constant</td>
</tr>
<tr>
<td>age</td>
<td>Constant</td>
</tr>
<tr>
<td>sex</td>
<td>Constant</td>
</tr>
</tbody>
</table>
Cox Model for E1684 Data: Uniform Priors

The PHREG Procedure

Bayesian Analysis

Initial Values of the Chains

<table>
<thead>
<tr>
<th>Chain</th>
<th>Seed</th>
<th>trt</th>
<th>age</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>532</td>
<td>-0.3598</td>
<td>0.00491</td>
<td>-0.0180</td>
</tr>
<tr>
<td>2</td>
<td>-0.7909</td>
<td>-0.0110</td>
<td>-0.4586</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0713</td>
<td>0.0209</td>
<td>0.4226</td>
<td></td>
</tr>
</tbody>
</table>

Fit Statistics

- AIC (smaller is better): 2020.131
- BIC (smaller is better): 2029.965
- DIC (smaller is better): 2020.154
- pD (Effective Number of Parameters): 3.012
Cox Model for E1684 Data: Uniform Priors

The PHREG Procedure

Bayesian Analysis

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>20000</td>
<td>-0.3608</td>
<td>0.1448</td>
<td>-0.4570</td>
<td>-0.3591</td>
<td>-0.2633</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>0.00489</td>
<td>0.00530</td>
<td>0.00136</td>
<td>0.00492</td>
<td>0.00846</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>-0.0210</td>
<td>0.1476</td>
<td>-0.1190</td>
<td>-0.0191</td>
<td>0.0783</td>
</tr>
</tbody>
</table>

Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.050</td>
<td>-0.6491</td>
<td>-0.0789</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.00569</td>
<td>0.0151</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.3148</td>
<td>0.2665</td>
</tr>
</tbody>
</table>
Cox Model with Normal Prior

The PHREG Procedure

Bayesian Analysis

Model Information

Data Set WORK.MELANOMA

Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>1</td>
<td>-0.3598</td>
<td>0.1437</td>
<td>-0.6415 -0.0782</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>0.00491</td>
<td>0.00532</td>
<td>-0.00551 0.0153</td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>-0.0180</td>
<td>0.1469</td>
<td>-0.3059 0.2698</td>
</tr>
</tbody>
</table>

Independent Normal Prior for Regression Coefficients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0</td>
<td>1E-6</td>
</tr>
<tr>
<td>age</td>
<td>0</td>
<td>1E-6</td>
</tr>
<tr>
<td>sex</td>
<td>0</td>
<td>1E-6</td>
</tr>
</tbody>
</table>
The PHREG Procedure

Bayesian Analysis

Initial Values of the Chains

<table>
<thead>
<tr>
<th>Chain</th>
<th>Seed</th>
<th>trt</th>
<th>age</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5432</td>
<td>-0.3598</td>
<td>0.00491</td>
<td>-0.0180</td>
</tr>
<tr>
<td>2</td>
<td>-0.7909</td>
<td>-0.0110</td>
<td>-0.4586</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0713</td>
<td>0.0209</td>
<td>0.4226</td>
<td></td>
</tr>
</tbody>
</table>

Fit Statistics

AIC (smaller is better) 2020.131
BIC (smaller is better) 2029.965
DIC (smaller is better) 2020.163
pD (Effective Number of Parameters) 3.016
Cox Model with Normal Prior

The PHREG Procedure
Bayesian Analysis
Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>20000</td>
<td>-0.3599</td>
<td>0.1451</td>
<td>-0.4561</td>
<td>-0.3583</td>
<td>-0.2612</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>0.00490</td>
<td>0.00533</td>
<td>0.00133</td>
<td>0.00494</td>
<td>0.00845</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>-0.0212</td>
<td>0.1468</td>
<td>-0.1186</td>
<td>-0.0208</td>
<td>0.0772</td>
</tr>
</tbody>
</table>

Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.050</td>
<td>-0.6464 -0.0791</td>
<td>-0.6508 -0.0841</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.00556 0.0154</td>
<td>-0.00535 0.0156</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.3119 0.2664</td>
<td>-0.3088 0.2685</td>
</tr>
</tbody>
</table>
### Piecewise Constant Hazards Model

#### Constant Hazard Time Intervals

<table>
<thead>
<tr>
<th>Interval</th>
<th>N</th>
<th>Event</th>
<th>Hazard Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0, 0.11233)</td>
<td>16</td>
<td>15</td>
<td>Lambda1</td>
</tr>
<tr>
<td>[0.11233, 0.150685)</td>
<td>14</td>
<td>14</td>
<td>Lambda2</td>
</tr>
<tr>
<td>[0.150685, 0.20685)</td>
<td>12</td>
<td>12</td>
<td>Lambda3</td>
</tr>
<tr>
<td>[0.20685, 0.26301)</td>
<td>12</td>
<td>12</td>
<td>Lambda4</td>
</tr>
<tr>
<td>[0.26301, 0.3274)</td>
<td>13</td>
<td>13</td>
<td>Lambda5</td>
</tr>
<tr>
<td>[0.3274, 0.42603)</td>
<td>13</td>
<td>13</td>
<td>Lambda6</td>
</tr>
<tr>
<td>[0.42603, 0.49315)</td>
<td>13</td>
<td>13</td>
<td>Lambda7</td>
</tr>
<tr>
<td>[0.49315, 0.62603)</td>
<td>14</td>
<td>13</td>
<td>Lambda8</td>
</tr>
<tr>
<td>[0.62603, 0.87945)</td>
<td>13</td>
<td>13</td>
<td>Lambda9</td>
</tr>
<tr>
<td>[0.87945, 1.01918)</td>
<td>13</td>
<td>13</td>
<td>Lambda10</td>
</tr>
<tr>
<td>[1.01918, 1.35479)</td>
<td>13</td>
<td>13</td>
<td>Lambda11</td>
</tr>
<tr>
<td>[1.35479, 1.7274)</td>
<td>13</td>
<td>13</td>
<td>Lambda12</td>
</tr>
<tr>
<td>[1.7274, 2.356165)</td>
<td>16</td>
<td>13</td>
<td>Lambda13</td>
</tr>
<tr>
<td>[2.356165, 3.241095)</td>
<td>14</td>
<td>13</td>
<td>Lambda14</td>
</tr>
<tr>
<td>[3.241095, Infty)</td>
<td>96</td>
<td>13</td>
<td>Lambda15</td>
</tr>
</tbody>
</table>
### Piecewise Constant Hazards Model

The PHREG Procedure

Bayesian Analysis

**Maximum Likelihood Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda1</td>
<td>1</td>
<td>0.4426</td>
<td>0.1658</td>
<td>0.1177 0.7675</td>
</tr>
<tr>
<td>Lambda2</td>
<td>1</td>
<td>1.2935</td>
<td>0.4923</td>
<td>0.3286 2.2585</td>
</tr>
<tr>
<td>Lambda3</td>
<td>1</td>
<td>0.8016</td>
<td>0.3177</td>
<td>0.1790 1.4243</td>
</tr>
<tr>
<td>Lambda4</td>
<td>1</td>
<td>0.8374</td>
<td>0.3319</td>
<td>0.1870 1.4879</td>
</tr>
<tr>
<td>Lambda5</td>
<td>1</td>
<td>0.8423</td>
<td>0.3285</td>
<td>0.1985 1.4862</td>
</tr>
<tr>
<td>Lambda6</td>
<td>1</td>
<td>0.5903</td>
<td>0.2309</td>
<td>0.1377 1.0429</td>
</tr>
<tr>
<td>Lambda7</td>
<td>1</td>
<td>0.9236</td>
<td>0.3607</td>
<td>0.2167 1.6305</td>
</tr>
<tr>
<td>Lambda8</td>
<td>1</td>
<td>0.5012</td>
<td>0.1958</td>
<td>0.1175 0.8848</td>
</tr>
<tr>
<td>Lambda9</td>
<td>1</td>
<td>0.2817</td>
<td>0.1101</td>
<td>0.0659 0.4976</td>
</tr>
<tr>
<td>Lambda10</td>
<td>1</td>
<td>0.5470</td>
<td>0.2134</td>
<td>0.1287 0.9653</td>
</tr>
<tr>
<td>Lambda11</td>
<td>1</td>
<td>0.2539</td>
<td>0.0985</td>
<td>0.0609 0.4470</td>
</tr>
<tr>
<td>Lambda12</td>
<td>1</td>
<td>0.2455</td>
<td>0.0948</td>
<td>0.0597 0.4314</td>
</tr>
<tr>
<td>Lambda13</td>
<td>1</td>
<td>0.1681</td>
<td>0.0648</td>
<td>0.0410 0.2952</td>
</tr>
<tr>
<td>Lambda14</td>
<td>1</td>
<td>0.1346</td>
<td>0.0521</td>
<td>0.0326 0.2367</td>
</tr>
<tr>
<td>Lambda15</td>
<td>1</td>
<td>0.0406</td>
<td>0.0156</td>
<td>0.0101 0.0711</td>
</tr>
<tr>
<td>trt</td>
<td>1</td>
<td>-0.3616</td>
<td>0.1437</td>
<td>-0.6432 -0.0799</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>0.00530</td>
<td>0.00532</td>
<td>-0.00512 0.0157</td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>-0.0188</td>
<td>0.1468</td>
<td>-0.3066 0.2690</td>
</tr>
</tbody>
</table>
# Piecewise Constant Hazards Model

## Independent Gamma Prior for Hazards

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Shape</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda1</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda2</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda3</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda4</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda5</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda6</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda7</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda8</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda9</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda10</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda11</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda12</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda13</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda14</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lambda15</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Piecewise Constant Hazards Model

The PHREG Procedure
Bayesian Analysis
Independent Normal Prior for Regression Coefficients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0</td>
<td>1E-6</td>
</tr>
<tr>
<td>age</td>
<td>0</td>
<td>1E-6</td>
</tr>
<tr>
<td>sex</td>
<td>0</td>
<td>1E-6</td>
</tr>
</tbody>
</table>

Initial Values of the Chains

<table>
<thead>
<tr>
<th>Chain</th>
<th>Seed</th>
<th>Lambda1</th>
<th>Lambda2</th>
<th>Lambda3</th>
<th>Lambda4</th>
<th>Lambda5</th>
<th>Lambda6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5432</td>
<td>0.4426</td>
<td>1.2935</td>
<td>0.8016</td>
<td>0.8375</td>
<td>0.8423</td>
<td>0.5903</td>
</tr>
<tr>
<td>2</td>
<td>0.1439</td>
<td>0.4130</td>
<td>0.2442</td>
<td>0.2551</td>
<td>0.2615</td>
<td>0.1826</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.3613</td>
<td>4.0519</td>
<td>2.6320</td>
<td>2.7496</td>
<td>2.7138</td>
<td>1.9089</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chain</th>
<th>Lambda7</th>
<th>Lambda8</th>
<th>Lambda9</th>
<th>Lambda10</th>
<th>Lambda11</th>
<th>Lambda12</th>
<th>Lambda13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9236</td>
<td>0.5012</td>
<td>0.2817</td>
<td>0.5470</td>
<td>0.2539</td>
<td>0.2455</td>
<td>0.1681</td>
</tr>
<tr>
<td>2</td>
<td>0.2862</td>
<td>0.1553</td>
<td>0.0872</td>
<td>0.1697</td>
<td>0.0793</td>
<td>0.0771</td>
<td>0.0528</td>
</tr>
<tr>
<td>3</td>
<td>2.9804</td>
<td>1.6177</td>
<td>0.9103</td>
<td>1.7632</td>
<td>0.8130</td>
<td>0.7822</td>
<td>0.5348</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chain</th>
<th>Lambda14</th>
<th>Lambda15</th>
<th>trt</th>
<th>age</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1346</td>
<td>0.0406</td>
<td>-0.3616</td>
<td>0.00530</td>
<td>-0.0188</td>
</tr>
<tr>
<td>2</td>
<td>0.0422</td>
<td>0.0128</td>
<td>-0.7926</td>
<td>-0.0106</td>
<td>-0.4593</td>
</tr>
<tr>
<td>3</td>
<td>0.4295</td>
<td>0.1282</td>
<td>0.0695</td>
<td>0.0212</td>
<td>0.4217</td>
</tr>
</tbody>
</table>
### Bayesian Analysis

#### Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda1</td>
<td>20000</td>
<td>0.4536</td>
<td>0.1742</td>
<td>0.3291</td>
<td>0.4266</td>
<td>0.5460</td>
</tr>
<tr>
<td>Lambda2</td>
<td>20000</td>
<td>1.3299</td>
<td>0.5167</td>
<td>0.9553</td>
<td>1.2521</td>
<td>1.6084</td>
</tr>
<tr>
<td>Lambda3</td>
<td>20000</td>
<td>0.8221</td>
<td>0.3338</td>
<td>0.5836</td>
<td>0.7665</td>
<td>0.9970</td>
</tr>
<tr>
<td>Lambda4</td>
<td>20000</td>
<td>0.8609</td>
<td>0.3480</td>
<td>0.6109</td>
<td>0.8074</td>
<td>1.0466</td>
</tr>
<tr>
<td>Lambda5</td>
<td>20000</td>
<td>0.8676</td>
<td>0.3520</td>
<td>0.6164</td>
<td>0.8059</td>
<td>1.0542</td>
</tr>
<tr>
<td>Lambda6</td>
<td>20000</td>
<td>0.6085</td>
<td>0.2443</td>
<td>0.4336</td>
<td>0.5680</td>
<td>0.7385</td>
</tr>
<tr>
<td>Lambda7</td>
<td>20000</td>
<td>0.9497</td>
<td>0.3800</td>
<td>0.6742</td>
<td>0.8915</td>
<td>1.1528</td>
</tr>
<tr>
<td>Lambda8</td>
<td>20000</td>
<td>0.5148</td>
<td>0.2064</td>
<td>0.3680</td>
<td>0.4814</td>
<td>0.6237</td>
</tr>
<tr>
<td>Lambda9</td>
<td>20000</td>
<td>0.2901</td>
<td>0.1174</td>
<td>0.2055</td>
<td>0.2705</td>
<td>0.3540</td>
</tr>
<tr>
<td>Lambda10</td>
<td>20000</td>
<td>0.5618</td>
<td>0.2229</td>
<td>0.4002</td>
<td>0.5262</td>
<td>0.6816</td>
</tr>
<tr>
<td>Lambda11</td>
<td>20000</td>
<td>0.2615</td>
<td>0.1043</td>
<td>0.1866</td>
<td>0.2442</td>
<td>0.3172</td>
</tr>
<tr>
<td>Lambda12</td>
<td>20000</td>
<td>0.2514</td>
<td>0.1002</td>
<td>0.1798</td>
<td>0.2356</td>
<td>0.3050</td>
</tr>
<tr>
<td>Lambda13</td>
<td>20000</td>
<td>0.1726</td>
<td>0.0681</td>
<td>0.1235</td>
<td>0.1617</td>
<td>0.2095</td>
</tr>
<tr>
<td>Lambda14</td>
<td>20000</td>
<td>0.1378</td>
<td>0.0547</td>
<td>0.0985</td>
<td>0.1293</td>
<td>0.1669</td>
</tr>
<tr>
<td>Lambda15</td>
<td>20000</td>
<td>0.0416</td>
<td>0.0162</td>
<td>0.0301</td>
<td>0.0391</td>
<td>0.0503</td>
</tr>
<tr>
<td>trt</td>
<td>20000</td>
<td>-0.3606</td>
<td>0.1447</td>
<td>-0.4576</td>
<td>-0.3589</td>
<td>-0.2652</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>0.00535</td>
<td>0.00534</td>
<td>0.00174</td>
<td>0.00526</td>
<td>0.00895</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>-0.0198</td>
<td>0.1482</td>
<td>-0.1195</td>
<td>-0.0187</td>
<td>0.0793</td>
</tr>
</tbody>
</table>
### Piecewise Constant Hazards Model

#### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda1</td>
<td>0.050</td>
<td>0.1993</td>
<td>0.1737</td>
</tr>
<tr>
<td>Lambda2</td>
<td>0.050</td>
<td>0.5691</td>
<td>0.4852</td>
</tr>
<tr>
<td>Lambda3</td>
<td>0.050</td>
<td>0.3382</td>
<td>0.2681</td>
</tr>
<tr>
<td>Lambda4</td>
<td>0.050</td>
<td>0.3508</td>
<td>0.2911</td>
</tr>
<tr>
<td>Lambda5</td>
<td>0.050</td>
<td>0.3572</td>
<td>0.2871</td>
</tr>
<tr>
<td>Lambda6</td>
<td>0.050</td>
<td>0.2543</td>
<td>0.2091</td>
</tr>
<tr>
<td>Lambda7</td>
<td>0.050</td>
<td>0.3956</td>
<td>0.3240</td>
</tr>
<tr>
<td>Lambda8</td>
<td>0.050</td>
<td>0.2124</td>
<td>0.1808</td>
</tr>
<tr>
<td>Lambda9</td>
<td>0.050</td>
<td>0.1208</td>
<td>0.1043</td>
</tr>
<tr>
<td>Lambda10</td>
<td>0.050</td>
<td>0.2337</td>
<td>0.1903</td>
</tr>
<tr>
<td>Lambda11</td>
<td>0.050</td>
<td>0.1100</td>
<td>0.0913</td>
</tr>
<tr>
<td>Lambda12</td>
<td>0.050</td>
<td>0.1049</td>
<td>0.0852</td>
</tr>
<tr>
<td>Lambda13</td>
<td>0.050</td>
<td>0.0728</td>
<td>0.0572</td>
</tr>
<tr>
<td>Lambda14</td>
<td>0.050</td>
<td>0.0566</td>
<td>0.0464</td>
</tr>
<tr>
<td>Lambda15</td>
<td>0.050</td>
<td>0.0176</td>
<td>0.0143</td>
</tr>
<tr>
<td>trt</td>
<td>0.050</td>
<td>-0.6465</td>
<td>-0.6471</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.00494</td>
<td>0.00492</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.3126</td>
<td>-0.3053</td>
</tr>
</tbody>
</table>
Further Prior Constructions in PHREG

data NormalPrior2;
  input _type_ $ _name_:$9. trt age sex;
cards;
  Mean . 1 2 3
  Cov trt 1 3 1
  Cov age 3 1 4
  Cov sex 1 4 1
;
run;

proc phreg data=melanoma;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
  coeffprior = normal(input=NormalPrior2)
run;
Further Prior Constructions in PHREG

data lambdaprior1;
input _type_ $ lambda1-lambda5;
cards; shape a1 a2 a3 a4 a5
iscale b1 b2 b3 b4 b5;
run;

proc phreg data=melanoma;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
coeffprior = normal(input=NormalPrior1)
  piecewise = hazard(prior=gamma(input=lambdaprior1));
run;
PHREG: Posterior Hazard Ratios

- PHREG enables the user to compute hazard ratios for class variables for the Cox or piecewise exponential models using the `hazardratio` statement.
- If A is a class variable, then we can use the statement

  ```
  hazardratio A;
  ```

- If X is a continuous variable, the following specification displays a table of hazard ratios comparing the hazards of each pair of levels of A at X=3.

  ```
  hazardratio A/ at (X = 3) diff = ALL;
  ```
E1690 Data

- E1690 was a follow-up confirmatory study to E1684 using the exact same patient population and same treatment arms. There were a total of $n = 427$ subjects on the observation and treatment arms combined for E1690.
- The primary endpoint was RFS.
Here, we can do a Bayesian analysis of E1690 in one of several ways:

- analyze the E1690 data alone with a noninformative prior on $\beta$.
- analyze the E1690 data and treat E1684 as historical data and construct a prior from the E1684 data using power prior ideas. We specify a prior of the form $\beta \sim N(\tilde{\beta}, a_0^{-1}\tilde{\Sigma})$, where $\tilde{\beta}$ is the posterior mean of $\beta$ based on a Bayesian analysis of E1684 using the Cox model with a uniform improper prior and $\tilde{\Sigma}$ is the posterior covariance matrix. We took $a_0 = 1$ in the analysis below.
- Combine both data sets into one big data set and introduce a binary covariate for study, and then do a Bayesian analysis on the combined data set using noninformative priors for $\beta$. 
data melanoma;
    infile 'H:/e1690ws.dat';
    input case study age trt sex perform nodes breslow failtime rfscens survtime scens;
run;

proc phreg;
    model failtime*rfscens(0) = trt age sex;
    bayes seed = 532 nbi=3000 nmc = 20000
        coeffprior = uniform
        summary = all
        diagnostics = none plots=none;
run;
SAS Code for E1690 Bayesian Analysis

data Normalprior2;
  input _type_ $ _name_:3. trt age sex;
cards;
    Mean .   -0.3608   .00489  - .0210
    Cov trt  2.096704e-02 -3.530224e-06 1.175486e-03
    Cov age  -3.530224e-06 2.809000e-05 4.028742e-05
    Cov sex  1.175486e-03 4.028742e-05 2.178576e-02
; run;

proc phreg;
  model failtime*rfscens(0) = trt age sex;
  bayes seed = 5432 nbi=3000 nmc = 20000
  coeffprior = normal(input=NormalPrior2)
  summary = all
  diagnostics = none plots=none;
run;
### Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>20000</td>
<td>-0.2136</td>
<td>0.1298</td>
<td>-0.3008</td>
<td>-0.2118</td>
<td>-0.1266</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>0.00880</td>
<td>0.00500</td>
<td>0.00548</td>
<td>0.00881</td>
<td>0.0122</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>-0.1472</td>
<td>0.1367</td>
<td>-0.2363</td>
<td>-0.1449</td>
<td>-0.0551</td>
</tr>
</tbody>
</table>

### Posterior Intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.050</td>
<td>-0.4729 0.0405</td>
<td>-0.4826 0.0294</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.00097 0.0186</td>
<td>-0.00097 0.0186</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.4232 0.1176</td>
<td>-0.4113 0.1259</td>
</tr>
</tbody>
</table>
Normal Prior for Regression Coefficients

Covariance Matrix

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>trt</th>
<th>age</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>-0.3608</td>
<td>0.020967</td>
<td>-3.53E-6</td>
<td>0.001175</td>
</tr>
<tr>
<td>age</td>
<td>0.00489</td>
<td>-3.53E-6</td>
<td>0.000028</td>
<td>0.00004</td>
</tr>
<tr>
<td>sex</td>
<td>-0.021</td>
<td>0.001175</td>
<td>0.00004</td>
<td>0.021786</td>
</tr>
</tbody>
</table>

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>20000</td>
<td>-0.2857</td>
<td>0.0969</td>
<td>-0.3508</td>
<td>-0.2855</td>
<td>-0.2204</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>0.00685</td>
<td>0.00362</td>
<td>0.00443</td>
<td>0.00687</td>
<td>0.00928</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>-0.0864</td>
<td>0.1000</td>
<td>-0.1532</td>
<td>-0.0860</td>
<td>-0.0192</td>
</tr>
</tbody>
</table>
### E1690: Cox model and Informative Normal Prior for $\beta$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.050</td>
<td>-0.4753</td>
<td>-0.4791</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>-0.00029</td>
<td>0.0139</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.2860</td>
<td>0.1092</td>
</tr>
</tbody>
</table>
Bayesian Analysis

Posterior Summaries

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>20000</td>
<td>-0.3062</td>
<td>0.0803</td>
<td>-0.3599</td>
<td>-0.3055</td>
<td>-0.2522</td>
</tr>
<tr>
<td>age</td>
<td>20000</td>
<td>0.00596</td>
<td>0.00300</td>
<td>0.00395</td>
<td>0.00596</td>
<td>0.00797</td>
</tr>
<tr>
<td>sex</td>
<td>20000</td>
<td>-0.0615</td>
<td>0.0824</td>
<td>-0.1166</td>
<td>-0.0611</td>
<td>-0.00559</td>
</tr>
</tbody>
</table>
### E1690: Cox model Based on Combined Datasets

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.050</td>
<td>-0.4637</td>
<td>-0.1494</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.4647</td>
<td>-0.1509</td>
</tr>
<tr>
<td>age</td>
<td>0.050</td>
<td>0.000103</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000111</td>
<td>0.0119</td>
</tr>
<tr>
<td>sex</td>
<td>0.050</td>
<td>-0.2241</td>
<td>0.0985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.2205</td>
<td>0.1008</td>
</tr>
</tbody>
</table>
Hazard Ratios for E1690

```
proc phreg;
  class trt;
  model failtime*rfscens(0) = trt age sex;
bayes seed = 532 nbi=3000 nmc = 20000
  coeffprior = uniform
  summary = all
  diagnostics = none plots=none;
hazardratio trt;
run;
```
Hazard Ratios for E1690

Hazard Ratios for trt

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt 0 vs 1</td>
<td>20000</td>
<td>1.2483</td>
<td>0.1633</td>
<td>1.1353</td>
<td>1.2373</td>
<td>1.3517</td>
</tr>
</tbody>
</table>

Hazard Ratios for trt

<table>
<thead>
<tr>
<th>95% Equal-Tail Interval</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9573 1.5978</td>
<td>0.9318 1.5654</td>
</tr>
</tbody>
</table>
Closing Remarks

- SAS has developed some excellent software for carrying out Bayesian analyses of generalized linear models and survival models.
- The software can handle large data sets and has a large suite of models and priors to choose from.
- The survival models are discussed in more detail in the book by Ibrahim, Chen, and Sinha (2001), titled *Bayesian Survival Analysis*, Springer-Verlag.
- There will be plenty of enhancements to come in future editions.
- Thanks to SAS for taking on this very important project!
References