



Getting Started with Bayesian Analysis

Danny Modlin

Senior Analytic Training Consultant
SAS – Cary Offices



Copyright © SAS Institute Inc. All rights reserved.

Objectives

- Introduce the basic concepts of Bayesian analysis.
- Illustrate the differences between Bayesian analysis and classical statistics.
- Introduce the Markov chain diagnostic statistics.
- Explain the advantages and disadvantages of Bayesian analysis.

What Is Bayesian Analysis?

- *Bayesian analysis* is a field of statistics that is based on the notion of conditional probability.
- It can be viewed as the formalization of the process of incorporating scientific knowledge using probabilistic tools.
- It provides uncertainty quantification of parameters by its conditional distribution in the light of available data.

3

Copyright © SAS Institute Inc. All rights reserved.



Bayes' Theorem

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

- $P(A)$ is the prior probability of event A. It is called the *prior* because it does not take into account any information about event B.
- $P(B | A)$ is the conditional probability of event B given event A.
- $P(B)$ is the prior or marginal probability of event B.
- $P(A | B)$ is the conditional probability of event A given event B. It is called the posterior probability because it is derived from the specified value of event B.

4

Copyright © SAS Institute Inc. All rights reserved.



Frequentist Approach to Statistics

- *Classical methods* consider the parameters to be fixed but unknown.
- They do not enable you to make probability statements about parameters because they are fixed.
- They are based on probabilities that are only for observations given the unknown parameters.
- They are judged by how they perform in an infinite number of hypothetical repetitions of the experiments.

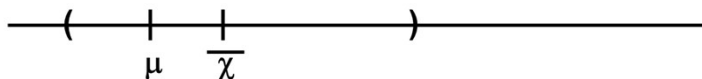
5

Copyright © SAS Institute Inc. All rights reserved.



Confidence Intervals – Classical Approach

95% Confidence



- A 95% confidence interval states that you are 95% confident that random interval contains the true mean.
 - In other words, if 100 different samples were drawn from the same population and 100 intervals were calculated, approximately 95 of them would contain the population mean.

6

Copyright © SAS Institute Inc. All rights reserved.



Bayesian Approach to Statistics

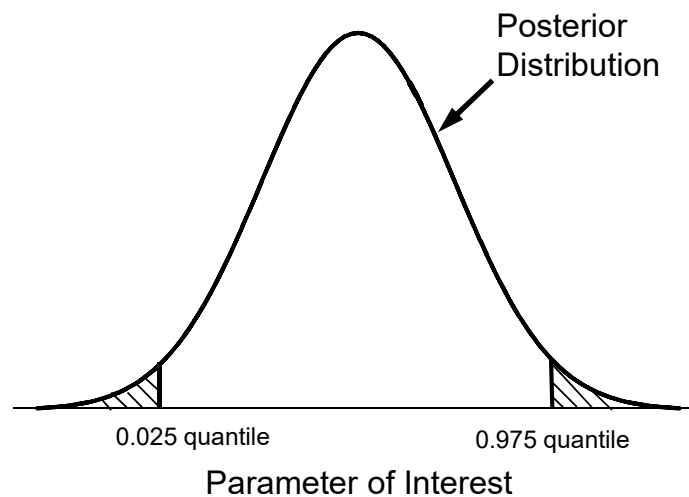
- Bayesian methods treat the unknown parameters as random variables, which enables you to make probability statements about them.
- Probabilities for parameters are interpreted as “degree of belief” and can be subjective.
- The rules of probability are used to revise “degree of beliefs” about the parameters given the observed data.
- The inferences about the parameters are based on the probability distribution for the parameter.

7

Copyright © SAS Institute Inc. All rights reserved.



95% Credible Interval



There is a 95% chance that the parameter is in the credible interval.

Copyright © SAS Institute Inc. All rights reserved.



Steps Involved in Bayesian Inference

1. The probability distribution of the parameter, known as the *prior distribution*, is formulated.
2. Given the observed data, you choose a statistical model (referred to as the likelihood) that describes the distribution of the data given the parameters.
3. You update your beliefs about the parameter by combining information from the prior distribution and the data through the calculation of the posterior distribution. This is carried out by using Bayes' theorem; hence the term Bayesian analysis.

9

Copyright © SAS Institute Inc. All rights reserved.



The Bayes' Rule

$$p(\theta | x) = \frac{f(x | \theta)\pi(\theta)}{m(x)}$$

Diagram illustrating the components of Bayes' Rule:

- posterior density of θ given x** points to $p(\theta | x)$.
- sampling density of x given θ** points to $f(x | \theta)$.
- prior density for θ** points to $\pi(\theta)$.
- marginal density of x** points to $m(x)$.

10

Copyright © SAS Institute Inc. All rights reserved.



Prior Distributions

- You cannot carry out any Bayesian inference or perform any modeling without using a prior distribution.
- It is not necessarily specified beforehand because prior does not refer to time.
- It is not necessarily unique, as the prior distribution could be a combination of prior distributions expressing a range of reasonable opinions.
- It is not necessarily completely specified, as it might be possible to have unknown parameters in the prior, which are then estimated.
- It is not necessarily important, as it could have a negligible influence on the conclusions, especially when the sample size is large.

11

Copyright © SAS Institute Inc. All rights reserved.



Noninformative vs Informative Priors

- A prior distribution is *noninformative* if it is flat, relative to the likelihood function. It will have minimal impact on the posterior distribution.
- An *informative prior* is a prior that is not strongly dominated by the likelihood and might have an impact on the posterior distribution.
 - The information can be obtained from the elicitation of expert opinion or the derivation from historical data.
 - It is recommended that you conduct a sensitivity analysis to assess the impact of a particular prior distribution on the conclusions of the analysis.

12

Copyright © SAS Institute Inc. All rights reserved.



Computational Issues

- The posterior distribution or any of its summary measures can only be obtained in closed form for a restricted set of relatively simple models.
- For many models, including generalized linear models, nonlinear models, random-effects models, and survival models, the posterior distribution does not have a closed form.
- In these situations, exact inference is not possible.

13

Copyright © SAS Institute Inc. All rights reserved.



Markov Chain Monte Carlo Methods

- *Monte Carlo methods* involve the use of random sampling techniques based on computer simulation to obtain approximate solutions to integration problems.
- *Markov Chain Monte Carlo (MCMC)* methods are an effective means of sampling from the posterior distribution of interest even when the posterior has no known closed algebraic form.
- Any inferences that you want to make about the parameters are derived from the sampled values.

14

Copyright © SAS Institute Inc. All rights reserved.



Markov Chain Convergence

- *Convergence* means that a Markov chain has reached its stationary (target) distribution.
- Assessing the Markov chain convergence is very important, as no valid inferences can be drawn if the chain is not converged.
- It is important to check the convergence for all the parameters and not just the ones of interest.
- Assessing convergence is a difficult task, as the chain converges to a distribution and not to a fixed point.

15

Copyright © SAS Institute Inc. All rights reserved.



Burn-In and Thinning

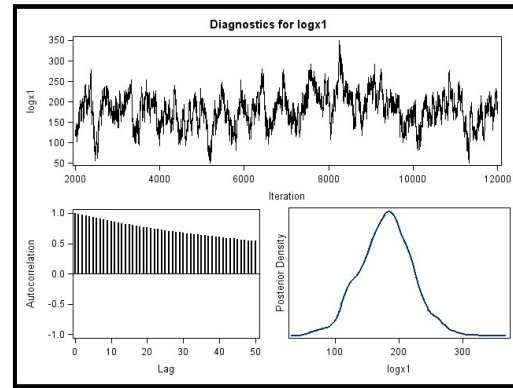
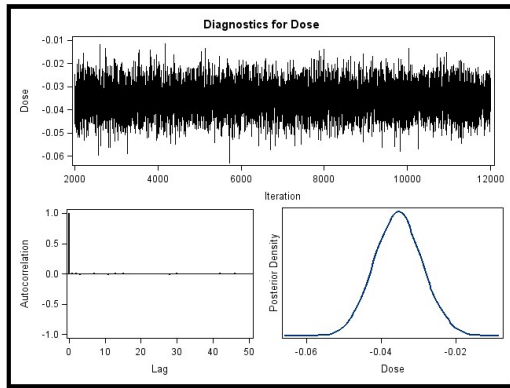
- *Burn-in* refers to the practice of discarding an initial portion of a Markov chain sample so that the effect of the initial values on the posterior inference is minimized.
- *Thinning* refers to the practice of keeping every k^{th} simulated draw from each sequence in order to reduce sample autocorrelations.
- Autocorrelations do not lead to biased Monte Carlo estimates, but rather it is an indicator of poor sampling efficiency.

16

Copyright © SAS Institute Inc. All rights reserved.



Diagnostic Plots – Good vs Poor Mixing



17

Copyright © SAS Institute Inc. All rights reserved.



Markov Chain Convergence Diagnostics

- Gelman and Rubin
- Geweke
- Heidelberger and Welch
- Raftery and Lewis
- Effective Sample Size

18

Copyright © SAS Institute Inc. All rights reserved.



Summary of Convergence Diagnostics

- There are no definitive tests of convergence.
- Visual inspection of the trace plots is often the most useful approach.
- Geweke and Heidelberger-Welch tests sometimes are statistically significant even when the trace plots look good.
- Oversensitivity to minor departures from stationarity does not impact inferences. Different convergence diagnostics are designed to protect you against different potential pitfalls.

19

Copyright © SAS Institute Inc. All rights reserved.



Deviance Information Criterion (DIC)

- *Deviance Information Criterion* (DIC) is a Bayesian alternative to AIC and BIC.
- It is a statistic where the smaller value indicates a better fit to the data set.
- DIC can be applied to non-nested models and models that have random effects.

20

Copyright © SAS Institute Inc. All rights reserved.



Advantages of Bayesian Analysis

- Bayesian analysis is useful when you have prior information, either expert opinion or historical knowledge, that you want to incorporate into the analysis.
- It is useful if you want to communicate your findings in terms of probability notions that can be more easily understood by non-statisticians.
- It provides inferences that are conditional on the data and are exact, without reliance on asymptotic approximation.
- It provides the full uncertainty of parameters via the posterior distribution in contrast to point estimates and standard errors only.
- The simulations make the computations tractable even for complex hierarchical models.

21

Copyright © SAS Institute Inc. All rights reserved.



Disadvantages of Bayesian Analysis

- It does not tell you how to select a prior and there is no one correct way to choose 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.
- It can produce posterior distributions that are heavily influenced by the priors.
- It often comes with a high computational cost, especially in models with a large number of parameters.

22

Copyright © SAS Institute Inc. All rights reserved.



Bayesian Analysis in SAS

Bayesian methods in SAS/STAT 15.1 are found in the following procedures:

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

23

Copyright © SAS Institute Inc. All rights reserved.



BAYES Statement

- The BAYES statement requests a Bayesian analysis of the regression model.
- The Bayesian posterior samples (also known as the chain) for the regression parameters can be output to a SAS data set.

```
proc genmod data=sasuser.birth desc;  
  model low=alcohol hist hyp mother wt prev preterm  
        / dist=binomial link=logit;  
  bayes seed=27513;  
  title 'Bayesian Analysis of Low Birth Weight Model';  
run;
```

24

Copyright © SAS Institute Inc. All rights reserved.



Simple Linear Regression

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

for subjects $i=1,2,\dots,n$,

The above model can equivalently be expressed as

$$Y_i \sim \text{normal}(\beta_0 + \beta_1 X_i, \sigma^2)$$

for subjects $i=1,2,\dots,n$

25

Copyright © SAS Institute Inc. All rights reserved.



Simple Linear Regression

```
proc mcmc data=slr seed=27513;
  parms beta0 0 beta1 0;
  parms sigma2 1;
  prior beta0 beta1 ~ normal(mean=0,
    var=1e6);
  prior sigma2 ~ igamma(shape=2.001,
    scale=1.001);
  mu=beta0 + beta1*X1;
  model Y ~ normal(mu, var=sigma2);
run;
```

26

Copyright © SAS Institute Inc. All rights reserved.



Ready to learn more?

Visit support.sas.com for the latest training and books.

SAS Training Course of Interest

Public training course

- Bayesian Analyses Using SAS



27

Copyright © SAS Institute Inc. All rights reserved.

Questions?

Copyright © SAS Institute Inc. All rights reserved.