

Paper 4745-2020

## Fitting Statistical Models with PROC MCMC

Mark Ghamsary, Keiji Oda, Larry Beeson, Loma Linda University.

## ABSTRACT

Bayesian inference, in particular Markov Chain Monte Carlo (MCMC), is one of the most important statistical tools for analyses. Although there is free access to many powerful statistical software tools for Bayesian analysis, still, it is challenging both to learn and to apply to real life research. SAS® has facilitated many procedures using Bayesian analysis which make it much easier to use, particularly for SAS users. This presentation **demonstrates various examples such as 'one sample proportion', 'two sample proportion', and 'two sample t-test', to more advanced models via Bayesian analysis. The results will be compared with non-Bayesian models. Many real-life examples in medicine, clinical trials and meta-analysis will be given.**

## Keywords

Gibbs Sampling, Markov Chain Monte Carlo, Meta-Analysis, Binary Regression

## INTRODUCTION

Despite differences in Bayesian and frequentist approaches we do have: Bernstein-von Mises Theorem: Under suitable assumptions and for sufficiently large sample sizes, the **posterior distribution of  $\theta$  is approximately normal with mean equal to the true value of  $\theta$**  and variance equal to the inverse of the Fisher information matrix. This theorem implies that Bayesian and Maximum Likelihood Estimate (MLE) estimators have the same large sample properties - not really surprising since influence of the prior should diminish with increasing sample sizes. But this is a **theoretical result and we often don't have "large"** sample sizes, so it is quite possible for the posterior to be (very) non-normal and even multi-modal. Most of Bayesian inference is concerned with (which often means simulating from) the posterior

$$f(\theta | y) = \frac{f(y | \theta) f(\theta)}{\int_{\theta} f(y | \theta) f(\theta) d\theta} \quad (1)$$

Clearly, there will be integration involved in (1). One of the important methods of integration is the acceptance-rejection algorithm, but this can work very well for low-dimensions, however it can be extremely inefficient for high dimensions. Although it can be a useful technique (even for high dimensions) when combined with MCMC methods.

MCMC is a combination of two terms, Markov Chain and Monte Carlo, where the second one **(Monte Carlo) was originally developed in the 1940's by** physicists at Los Alamos. Bayesians, and sometimes also frequentists, need to integrate over possibly high-dimensional probability distributions to make inference about model parameters or to make predictions. Bayesians need to integrate over the posterior distribution of the model

parameters given the data, and frequentists may need to integrate over the distribution of observables given the parameter values. As described below, *Monte Carlo* integration draws samples from the required distribution, and then forms sample averages to approximate expectations. *Markov Chain Monte Carlo* draws these samples by running a cleverly constructed Markov chain for a long time.

Metropolis et al. (1953) introduced a Monte Carlo-type algorithm to investigate the equilibrium properties of large systems of particles, such as molecules in a gas. Hastings (1970) used the Metropolis algorithm to sample from certain distributions; for example, normal (standard), Poisson, and random orthogonal matrices. Geman and Geman (1984) illustrated the use of a version of the algorithm that they called the Gibbs sampler in the context of image reconstruction. Tanner and Wang (1987) developed a framework in which Gibbs sampler algorithms can be used to calculate posterior distributions, they called it *Data Augmentation*. Gelfand, Hills, Racine-Poon, and Smith (1990), Gelfand and Smith (1990), and Zeger and Karim (1991) used the Gibbs sampler to perform Bayesian computation in various important statistical problems. Gibbs sampling is a special case of Markov Chain Monte Carlo (MCMC) using the Metropolis-Hastings algorithm, which is a general method for the simulation of stochastic processes having conditional probability densities known up to a constant of proportionality.

Before we go through PROC MCMC we need to explain how Gibbs sampling is working. We start by a simple example from Casella and George (1992) paper.

**Example1.** For the following joint distribution of  $\mathbf{x}$  and  $\mathbf{y}$

$$f(\mathbf{x}, \theta) \propto \binom{n}{x} \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}, \quad x = 0, 1, \dots, n, 0 \leq \theta \leq 1,$$

The full conditionals are the given by:

$$1. \quad f(x | \theta) \propto \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x},$$

$$\bullet (x | \theta) \sim \text{Bin}(n, \theta).$$

$$2. \quad f(\theta | x) \propto \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1},$$

$$\bullet (\theta | x) \sim \text{Beta}(x + \alpha, n - x + \beta).$$

We will go through this k=500 times and we will take the last values of x and theta as our first sample. We will repeat this N=5000 times.

Let us pick  $x^{(0)} = 5$ ,  $\theta^{(0)} = 0.20$ ,  $n = 10$  as initial values.

First cycle:

Sample  $x^{(1)}$  from  $(x^{(1)} | \theta^{(0)} = 0.50) \sim \text{Bin}(10, 0.20)$ ,

it gives:  $x^{(1)} = 0$ ,  $\theta^{(0)} = 0.20$ .

sample  $\theta^{(1)}$  from  $(\theta^{(1)} | x^{(1)} = 0) \sim \text{Beta}(0 + 1, 10 - 0 + 1)$ ,

$x^{(1)} = 0$ ,  $\theta^{(1)} = 0.01513$

Second cycle:

Sample  $x^{(2)}$  from  $(x^{(2)} | \theta^{(0)} = 0.01513) \sim \text{Bin}(10, 0.01513)$ ,

it gives:  $x^{(2)} = 2, \theta^{(1)} = 0.164732$ .

sample  $\theta^{(3)}$  from  $(\theta^{(3)} | x^{(2)} = 2) \sim \text{Beta}(2+1, 10-2+1)$ ,

$x^{(2)} = 2, \theta^{(1)} = 0.1666$ .

This can be repeated 500 times and pick the first sample. We will keep doing this from another 5000 times. The following SAS codes will do the job.

Beta Binomial
<pre> %let Sample = 5000; data BetaBin; alpha = 2; beta=4; n= 16; x=0 ;k=500;          /* initialize hyperparameters */ call streaminit(4321); do j=1 to &amp;Sample;;   do i = 1 to k;     a=x+alpha;b=n-x+beta;     x = rand("Binomial", tetha, n);/* x[i] tetha[i] ~ Bin(tetha[i], n) */     tetha = rand("Beta", a, b); /* tetha[i] x[i] ~ Beta(a,b) */     output;   end; end; keep x tetha; run; proc univariate data=betabin; var x tetha; histogram x tetha;run; </pre>

Table 1: SAS codes for Gibbs sampling of 2 variables  $x$  and  $\theta$ .

The output will be the distribution of both  $x$  and  $\theta$ . Then from the sample we can find any characteristic of the two random variables such as mean, median and standard deviation.

In SAS PROC MCMC Statements has the following syntax:

PROC MCMC options;

PARMS; /\* to declare the parameters  $\mu$  0;  $\sigma$  5

PRIOR; /\* Stating the distribution of priors. We can have more than one prior

abatelement  $\alpha \sim \text{Beta}(a, b), \beta \sim \text{Normal}(0, 100)$

Programming statements; /\*

MODE: like  $y \sim \text{Normal}(\mu, \sigma^2)$ , multiple model is allowed

RANDOM;  $\mu \sim \text{Normal}(10, 25)$ , multiple random statement is allowed

PREDDIST; Prediction distribution

Run;

## STEPS OF BAYESIAN METHOD

1. Decide on the **Prior**:  $P(\theta)$ 
  - $P(\theta)$  expresses what is known about  $\theta$  prior to observing  $\mathbf{y}$
2. Decide on the **Likelihood**:  $L(\bar{\mathbf{y}} | \theta)$ 
  - Describe the process of giving rise to the data in terms of unknown  $\theta$
3. Derive the **Posterior**:  $P(\theta | \bar{\mathbf{y}}) \propto L(\bar{\mathbf{y}} | \theta) \times P(\theta)$ 
  - Apply the Bayes Theorem to derive  $P(\theta | \bar{\mathbf{y}})$ , this expresses what is known about  $\theta$  after observing  $\mathbf{y}$ .
4. Inference statement are derived from **posterior** distribution  $P(\theta | \bar{\mathbf{y}})$ 
  - Point estimates, interval estimates, probability of hypotheses.
  - All parameters are assumed to be unknown

The diagram shows the Bayes' theorem formula:  $f(\theta | \mathbf{y}) = \frac{f(\mathbf{y} | \theta) f(\theta)}{\int_{\theta} f(\mathbf{y} | \theta) f(\theta) d\theta} \propto f(\mathbf{y} | \theta) f(\theta)$ . A red arrow points from the label 'Posterior' to  $f(\theta | \mathbf{y})$ . A green arrow points from 'Likelihood' to  $f(\mathbf{y} | \theta)$ . A blue arrow points from 'Prior' to  $f(\theta)$ . The entire equation is enclosed in a red rectangular box.

**Example2:** One sample Proportion

Let  $Y$  be a Bernoulli trial with probability of success  $\theta$ . Derive the posterior of  $\theta$  given some data.

**Likelihood:**  $Y | \theta \sim \text{Bin}(n, \theta)$ :  $L(\bar{\mathbf{y}} | \theta) = \binom{n}{\mathbf{y}} \theta^{\mathbf{y}} (1 - \theta)^{n - \mathbf{y}}$

**Prior:**  $\theta \sim \text{Beta}(a, b)$ :  $g(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha - 1} (1 - \theta)^{\beta - 1}$ ,

where,  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$ .

Posterior:  $f(\theta | y) = \frac{L(y | \theta)g(\theta)}{f(y)} = \frac{L(y | \theta)g(\theta)}{\int_0^1 L(y | \theta)g(\theta)d\theta}$

$$= \frac{\left[ \binom{n}{y} \theta^y (1-\theta)^{n-y} \right] \left[ \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \right]}{f(y)} = C \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1} \text{ where}$$

$$C = \left[ \binom{n}{y} \frac{1}{B(\alpha, \beta) f(y)} \right]$$

$$\therefore (\theta | y) \sim \text{Beta}(y + \alpha, n - y + \beta)$$

That is,  $(\theta | y)$  is a beta distribution with parameters  $y + \alpha$  and  $n - y + \beta$  respectively.

Let us use data: let  $Y \sim \text{Bin}(25, \theta)$  and we observed  $y = 12$  successes. Then we will get the posterior mean for four different priors as sensitivity of prior, Chen, F. (2009):

$$\hat{\theta}_{MLE} = \frac{12}{25} = 0.48$$

Prior	Posterior	Mean	Credible Interval
$\theta \sim \text{Beta}(1,1)$	$\theta   y \sim \text{Beta}(13,14)$	0.481	0.283 0.662
$\theta \sim \text{Beta}(3,7)$	$\theta   y \sim \text{Beta}(15,20)$	0.429	0.273 0.574
$\theta \sim \text{Beta}(7,3)$	$\theta   y \sim \text{Beta}(19,16)$	0.537	0.365 0.679
$\theta \sim \text{Beta}(10,10)$	$\theta   y \sim \text{Beta}(22,23)$	0.489	0.335 0.622

Table 2 the results on posterior of  $\theta$  for different distribution of priors.

In the following we use PROC MCMC where we used a Macro to calculate the posterior distribution of  $\theta$  for different prior distributions.

```

Proc MCMC
data one;
  n = 25; y = 12;
run;
%macro BetaBinom(alpha, beta);
ods graphics on;
title "Priors: Beta(&alpha, &beta)";
proc mcmc data=one seed=123 outpost=PosteriorSample diag=ess
  nbi=1000 nthin=10 nmc=10000 statistics=(summary interval);
  parms theta 0.5;
  prior theta ~ beta(&alpha, &beta);
  model y ~ binomial(n, theta);
  ods select PostSummaries PostIntervals TADPanel;
run;
ods graphics off;
%mend;
%BetaBinom(1, 1);
%BetaBinom(3, 7);
%BetaBinom(7, 3);
%BetaBinom(10, 10);

```

Table 3: SAS codes to produce the results founded in table 2

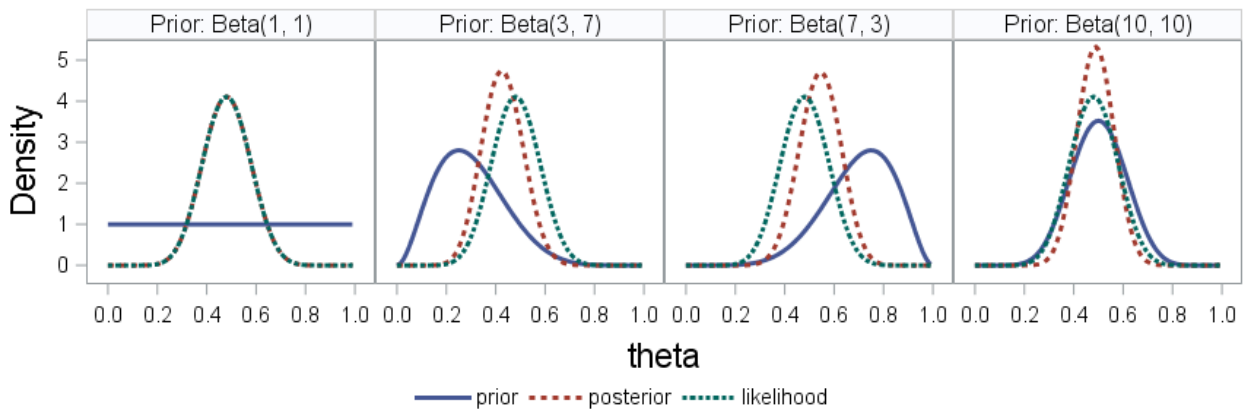


Figure 1: Distribution Plots for prior, posterior and likelihood of  $\theta$

The MCMC procedure produces the following three types of plots (Figure2) that can be used for convergence criterion. The first one on the top (trace plot) indicates that the Markov chain appears to be stable and constant variance. In other words, it is a good mixing. The second one (bottom left) is the autocorrelation plot that shows a very small degree of autocorrelation among the posterior samples for  $\theta$ . Finally, the third one is the kernel density plot estimates the posterior marginal distribution for the parameter  $\theta$ .

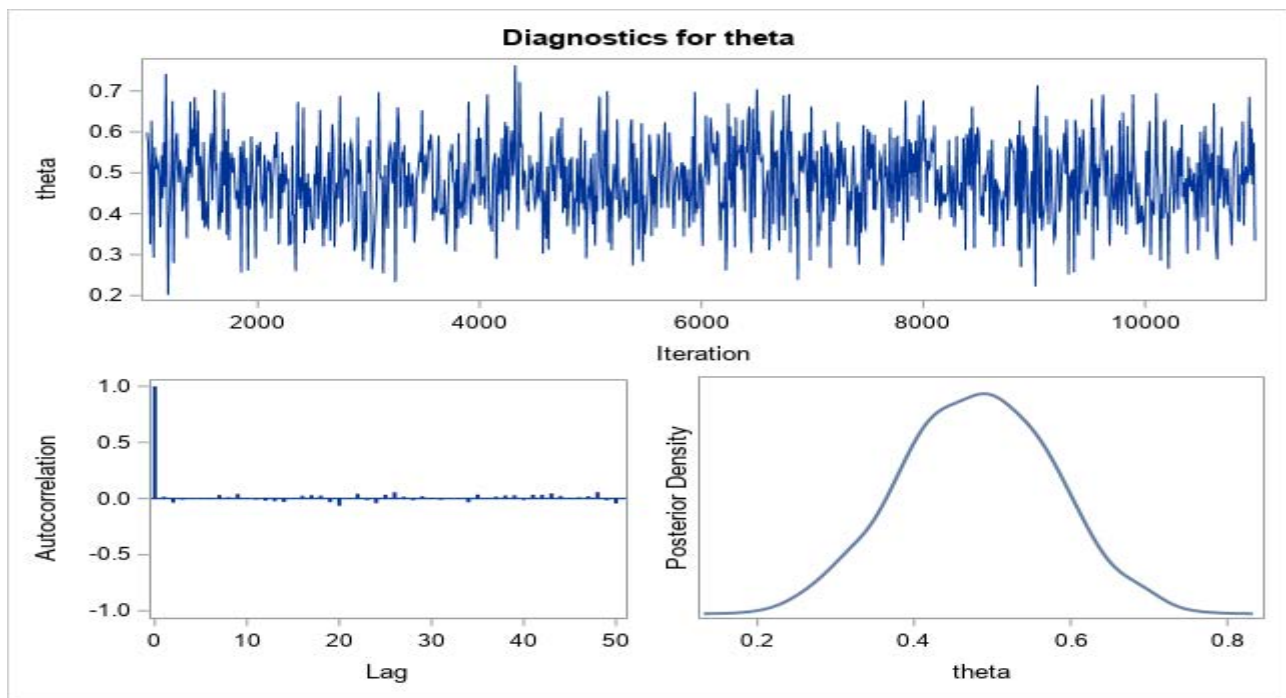


Figure 2 Diagnostic Plots for  $\theta$

## BINARY REGRESSION

Binary outcomes in many studies are usually analyzed via the logistic regression model to obtain odds ratios in order to compare two different levels of exposures. Recently, many papers have been published either with simulations or with real life data that are common (incidence of 10% or more); it is often more desirable to estimate a RR rather than OR with increasing incidence rates, and there is a tendency for some to interpret ORs as if they are RRs (McNutt LA, et al. 2003), but there is an increasing differential between the RR and OR when the incidence becomes more common.

For instance, let  $\hat{\theta}_1$  be proportion of say low birthweight children with smoking mothers and  $\hat{\theta}_2$  be proportion of low birthweight children among non-smoking mothers. There are three common measures of association in Epidemiology as follows:

- o Risk Difference (RD):  $\hat{\theta}_1 - \hat{\theta}_2$
- o Risk Ratio (RR):  $\hat{\theta}_1 / \hat{\theta}_2$
- o Odds Ratio (OR):  $\frac{\hat{\theta}_1 / (1 - \hat{\theta}_1)}{\hat{\theta}_2 / (1 - \hat{\theta}_2)}$

Example3: Risk Difference (RD): To make Bayesian inference on  $\hat{\theta}_1 - \hat{\theta}_2$ , we are going to apply a linear model to produce a binary outcome.

Model:

$$p_i = \beta_0 + \beta_1 X_i, i = 1, \dots, n,$$

with  $E(Y_i) = p_i$ , where  $Y_i = 1$  if Low Birth and 0 otherwise,  $X_i = 1$  if smoker and  $X_i = 0$  otherwise. Thus,

$$\begin{cases} \hat{\theta}_1 = E(Y_i = 1) = \hat{\beta}_0 + \hat{\beta}_1 \\ \hat{\theta}_2 = E(Y_i = 0) = \hat{\beta}_0 \end{cases} \Rightarrow \hat{\theta}_1 - \hat{\theta}_2 = \hat{\beta}_1$$

The biggest issue with binary regression is convergence. Pedroza and Truong (2016) used some simulations study in many cases and in all vases analyzed, it was observed that Bayesian methods are capable of estimating the measures of interest, always within the correct parametric space of probabilities. In particular we can use the multi-clinic data which is an easy approach via MMCMC. As a real life's example we use the data from a 1986 cohort study conducted at the Baystate Medical Center, Springfield Massachusetts, see Hosmer and Lemeshow (2000). The outcome is the low birthweight (1 if the weight is <2500 gr, 0 otherwise). The data consist of 189 live births of which 59 of them had low birth weight.

A frequentist approach is via Proc Freq to calculate the risk difference  $\hat{\theta}_1 - \hat{\theta}_2$  as shown here:

Risk difference						
<code>data one;set sasuser.lowirth;</code>						
<code>proc sort data=one; by descending low descending smoke;run;</code>						
<code>proc freq data=one order=data;</code>						
<code>tables smoke*low/nopercent nocol riskdiff all;run;</code>						
	Column 1 Risk Estimates					
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Difference	0.1532	0.0700	0.0161	0.2904		

Table 4: SAS codes and the partial output for risk differences  $\hat{\theta}_1 - \hat{\theta}_2$

$$\hat{\theta}_1 = \frac{30}{74} = 0.4054, \hat{\theta}_2 = \frac{29}{86} = 0.2522$$

$$\hat{\theta}_1 - \hat{\theta}_2 = 0.4052 - 0.2522 = 0.1532$$



Frequency Row Pct	Table of smoke by low			
	smoke	low		
		1	0	Total
<b>1</b>		30 40.54	44 59.46	74
<b>0</b>		29 25.22	86 74.78	115
<b>Total</b>		59	130	189

Now we will show binary regression with identity link as follow

```

Risk difference
proc genmod data=TWO desc;
model low=smoke;
bayes seed = 21 coeffprior = normal nmc = 10000 outpost =
mcmcout seed = 21 nbi = 1000 thin = 10;run;

```

Table 5: MCMC codes via PROC GENMOD for binary linear regression

$$\hat{\theta}_1 - \hat{\theta}_2 = \hat{\beta}_1 \quad \hat{y} = 0.2523 + 0.1532X_i$$

Parameter	N	Mean	Standard Deviation	Posterior Intervals			
				Equal-Tail Interval		HPD Interval	
Intercept	10000	0.2523	0.0432	0.1681	0.3363	0.1726	0.3403
SMOKE	10000	0.1531	0.0695	0.0158	0.2883	0.0175	0.2896

Output 1: MCMC output for,  $\hat{\beta}_1$  the regression coefficient

As it is shown in output 1, in above, there is no issue about the convergence. There is no apparent issues about the convergence. This has been also confirmed by the Geweke test (P-value=0.12) (Geweke (1992)). Also the effective sample size in all cases is above 9967 which is an indication of good mixing.

#### Example4: Risk Ratio (RR):

To estimate risk ratio (RR) we can use a log-binomial regression model, but it has again convergence issues. However, as we stated in above Bayesian simulations analysis, many have shown that this problem has been resolved by choosing an appropriate prior. Pedroza et al.(2016) in an extensive simulations study used the Bayesian approach to solve the problem of convergence.

Model:

$$\ln(p_i) = \beta_0 + \beta_1 X_i, \quad i = 1, \dots, n$$

where  $Y_i$  and  $X_i$  are defined as above. For the low birth data set we have:

$$\ln(p=1 | X=1) = \beta_0 + \beta_1 \Rightarrow \hat{\theta}_1 = \exp\{\beta_0 + \beta_1\}$$

$$\ln(p=1|X=0) = \beta_0 \Rightarrow \hat{\theta}_2 = \exp\{\beta_0\}$$

$$\frac{\hat{\theta}_1}{\hat{\theta}_2} = \frac{\exp\{\beta_0 + \beta_1\}}{\exp\{\beta_0\}} = e^{\beta_1}, \text{ as it is called risk ratio.}$$

Risk Ratio			
<pre> /* Log-Binomial*/ proc genmod data=one desc; model low=smoke/d=bin link=log; bayes seed = 1 coeffprior = normal; ods output PostSummaries = ps;run; data CI;   set ps (where = (Parameter ^= "Intercept"));   RR = exp(Mean);   Lower = exp(Mean - 1.96 * StdDev);   Upper = exp(Mean + 1.96 * StdDev);run; proc print;run; </pre>			
	Risk	95% Confidence Limits	
	1.59074	1.04016	2.43277

Table 6: SAS codes for MCMC codes via PROC GENMOD for log-binomial regression

$$\ln p = -0.9550 + 0.4999 X_i$$

$$\frac{\hat{\theta}_1}{\hat{\theta}_2} = \exp(\hat{\beta}_1) = \exp(0.4642) = 1.591$$

Parameter	N	Mean	Standard Deviation	95% Credible Interval			
				Equal-Tail Interval		HPD Interval	
Intercept	10000	-1.3958	0.1618	-1.7208	-1.0921	-1.7121	-1.0892
SMOKE	10000	0.4642	0.2167	0.0334	0.8764	0.0300	0.8677

Output 2: MCMC output for  $\hat{\beta}_1$ , the log-binomial regression coefficient

Again there is no apparent issue about the convergence. This has been also confirmed by graphical search and the Geweke test (P-value=0.47).

Example5: Odds Ratio (OR):

To estimate the OR, we can use a *logistic regression* as it has no serious issue with convergence, but it will not do well for high incidence (usually more than 10%). Here also Bayesian analysis has been used.

Model:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i.$$

For the low birth data set we have:

$$\ln\left(\frac{(p=1|X=1)}{1-(p=1|X=1)}\right) = \beta_0 + \beta_1 \Rightarrow \text{Odd}_1 = \exp\{\beta_0 + \beta_1\}$$

$$\ln\left(\frac{(p=1|X=0)}{1-(p=1|X=0)}\right) = \beta_0 \Rightarrow \text{Odd}_2 = \exp\{\beta_0\}$$

$$\ln(p=1|X=0) = \beta_0 \Rightarrow \hat{\theta}_2 = \exp\{\beta_0\}$$

$$\text{OR} = \frac{\exp\{\beta_0 + \beta_1\}}{\exp\{\beta_0\}} = e^{\beta_1}, \text{ as it is called odds ratio}$$

SAS Variable Format			
<pre> /* Logistic*/ proc genmod data=one desc; model low=smoke/d=bin link=logit; bayes seed = 1 coeffprior = normal; ods output PostSummaries = ps;run; data CI;   set ps (where = (Parameter ^= "Intercept"));   RR = exp(Mean);   Lower = exp(Mean - 1.96 * StdDev);   Upper = exp(Mean + 1.96 * StdDev);run; proc print;run; </pre>			
	OR	95% Confidence Limits	
	2.03459	1.07393	3.85459

Table 7: SAS codes for MCMC codes via PROC GENMOD for logistic regression

$$\ln\left(\frac{p_i}{1-p_i}\right) = -1.0968 + 0.7103 X_i, \quad \text{OR} = \exp(\hat{\beta}_1) = \exp(0.7103) = 2.03$$

Parameter	N	Mean	Standard Deviation	Posterior Intervals			
				Equal-Tail Interval		HPD Interval	
Intercept	10000	-1.0968	0.2153	-1.5218	-0.6865	-1.5105	-0.6800
SMOKE	10000	0.7103	0.3260	0.0601	1.3460	0.0747	1.3572

Output 2: MCMC output for  $\hat{\beta}_1$ , the -logit regression coefficient

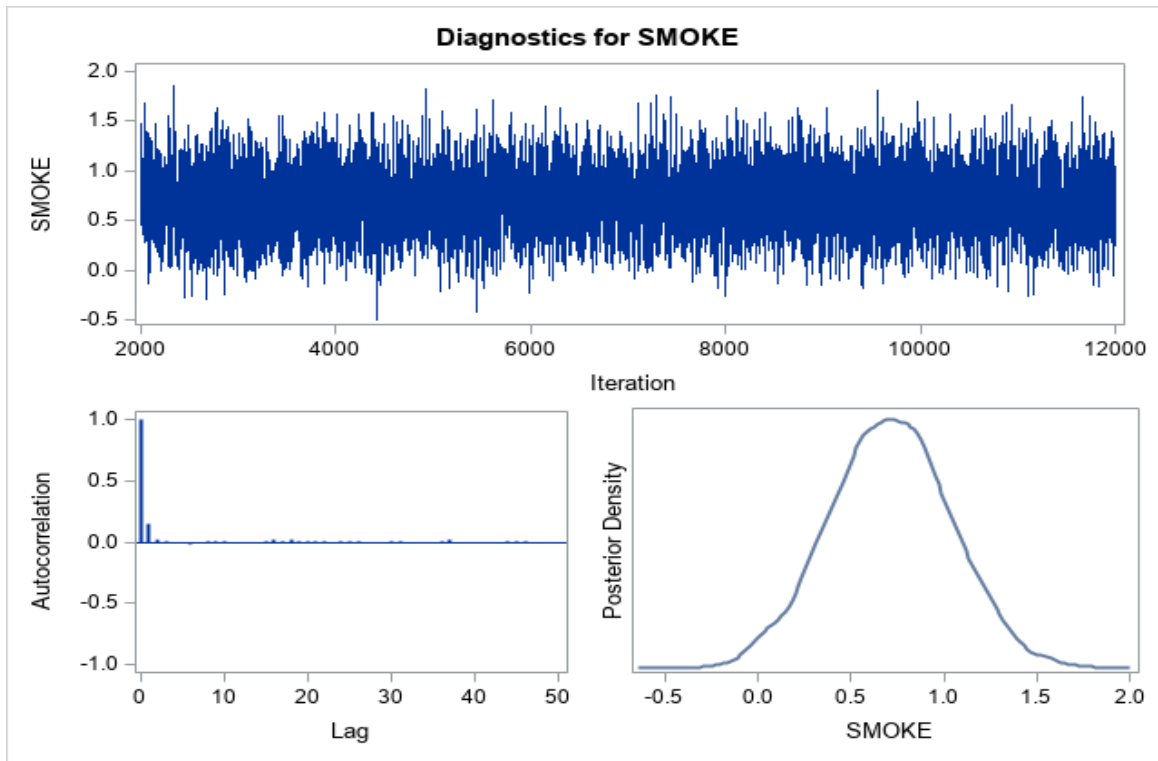


Figure 3 Diagnostic Plots for  $\beta_1$ , slope of log Odds Ratio on Smoking.

In Figure 3, in above, there is no obvious issue about the convergence and mixing. This has been also confirmed by the Geweke test (P-value=0.58).

**Remark:** All three binary models discussed above can easily be extended with a random effect model. For instance, let us try the effect of smoking in low-birth data from  $k$  different clinics with relative risk or RR as a measure of association. The model can be writes as follow:

$$\ln(p_{ij}) = \beta_0 + \beta_1 X_{ij} + \delta_j, \quad i = 1, \dots, n \quad \delta_j \sim N(0, \tau^2), \quad j = 1, 2, \dots, k.$$

For more details see Pedroza et al. (2017).

**Example6:** Two sample t-Test

For the two sample t-test it is the best to use a regression approach in PROC MCMC.

Model:

$$y_{ij} \sim N(\beta_0 + \beta_1 z_{ij}, \sigma_j^2); \quad i = 1, 2, \dots, n_{ij}, \quad j = 1, 2 \quad z_{ij} = \begin{cases} 1 & \text{if group 1} \\ 0 & \text{if group 2} \end{cases}$$

The frequentist approach assumes that the variances are equal and performs a two-sample t-test to get confidence intervals and test of hypothesis for the means. If the assumption fails then they use Satterwait- approximation as follows, but Casella and Berger (2002 page 410) claim the exact t-distribution in the case of unequal variances is not pleasant and the approximation may not work well and even result in incorrect conclusions due to the sensitivity of the test to the assumptions. It is always safe to assume they have unequal variance and in Bayesian analysis will treat the model as if they are unequal.

If  $\sigma_1 \neq \sigma_2$ , then use the following t statistic test.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Let us take a look at the example from Casella and Berger (2002 page 409):

Samples of wood are obtained from the corn and periphery of a certain Byzantine church. The data of the wood were determined, giving the following data

Corn	1294 1279 127 1264 1263 1254 1251 1251 1248 1240 1232 1220 1218 1210
Periphery	1284 1272 1256 1254 1242 1274 1264 1256 1250

Use two sample t-test to determine if the mean age is the same as the mean of the periphery.

```

Two sample T test

data p841; /* from casella and Burger Page409*/
input trt y@@;
cards;
1      1294      1      1251      1      1279      1      1248      1      1274      1      1240
1      1264      1      1232      1 1263      1      1220      1      1254      1      1218
1      1251      1      1210      2      1284      2      1274      2      1272      2      1264
2      1256      2      1256      2      1254      2      1250      2 1242
;
proc ttest data=p841;
class trt;
var y;run;
proc genmod data = SBP;
class trt;
model y = trt;
bayes seed = 1 coeffprior = normal nbi=1000 nmc=10000 thin=10 seed=140
out=posterior DIAGNOSTICS=ALL;
run;

```

Table 8: SAS codes for t test and MCMC codes via PROC GENMOD for the regression model

Frequentist Results								
	trt	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev		
	1		1249.9	1235.8	1263.9	24.3179	17.6294	39.1772
	2		1261.3	1251.1	1271.5	13.2665	8.9610	25.4156
	Diff (1-2)	Pooled	-11.4762	-29.9676	7.0152	20.8118	16.0116	29.7413
	Diff (1-2)	Satterthwaite	-11.4762	-27.8417	4.8893			

Bayesian Results							
Parameter	N	Mean	Standard Deviation	95% Credible Interval			
				Equal-Tail Interval		HPD Interval	
Intercept	10000	1261.3	7.2914	1246.6	1275.6	1246.7	1275.7
trt1	10000	-11.4722	9.3776	-30.0798	7.1527	-29.3786	7.6735

Output 3: the Bayesian result and frequentist output for two sample test.

As we observe the results both frequentist and Bayesian are pretty close in conclusion as expected since we used non-informative prior. We also checked for convergence and all assumptions were met.

**Example7: Bayesian simple linear regression:** In this example we will regress  $Y_i$ , the Systolic Blood Pressure (SBP), on  $X_i$ , age, a hypothetical data set. In this example, first we will try to do by the frequentist approach and then via MCMC, both by SAS. We will also add a random effect to the model and see how simple it can be done by PROC MCMC. The output for the random effect is not shown.

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, 20.$$

$$\begin{aligned} \text{Likelihood: } L(\bar{y} | \bar{x}, \beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} [y_i - (\beta_0 + \beta_1 x_i)]^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2\right\} \\ &\propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2\right\} \end{aligned}$$

$$\text{Prior: } h(\beta_0, \beta_1, \sigma) = h(\beta_0)h(\beta_1)h(\sigma) \propto \frac{1}{\sigma^2}$$

$$\text{Posterior: } P(\beta_0, \beta_1, \sigma^2 | \bar{y}, \bar{x}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2\right\}$$

Now we use the traditional Bayesian approach and compare with MCMC.

To get the marginal posteriors we integrate out as follows:

**Gibbs Sampling:** We can indirectly sample a random variable from a distribution without having the density (see Casella and George (1992)). This can be done by having conditional posterior densities as follows:

$$P(\beta_0, \beta_1, \sigma^2 | \vec{y}, \vec{x}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 \right\}$$

Let us define  $SSE = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$ , then we have:

$$(\beta_0, \beta_1 | \sigma^2, \mathbf{x}, \mathbf{y}) = N \left[ \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right]$$

$$\left( \frac{1}{\sigma^2} | \beta_0, \beta_1, \mathbf{x}, \mathbf{y} \right) = \text{Gamma} \left[ \frac{n-2}{2}, \frac{SSE}{2} \right]$$

**Informative Prior:** In case we do have informative prior as

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} | \sigma^2 \sim N \left[ \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \Sigma_0 \right]$$

$$\frac{1}{\sigma^2} | \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \text{Gamma} \left( \frac{k}{2}, \frac{k\sigma_0^2}{2} \right)$$

**Posterior:**

$$(\beta_0, \beta_1 | \sigma^2, \mathbf{x}, \mathbf{y}) = N \left[ \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \left( \Sigma_0 + \frac{1}{\sigma^2} (\mathbf{X}'\mathbf{X}) \right)^{-1} \right]$$

$$\frac{1}{\sigma^2} | \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \text{Gamma} \left( \frac{k+2}{2}, \frac{k\sigma_0^2 + SSE}{2} \right)$$

Simple Linear Regression							
<b>data</b>	one;						
<b>input</b>	ID	age	SBP	clin@@;			
<b>cards;</b>							
1	20	120	1	11	18	125	2
2	43	128	1	12	42	122	2
3	63	141	1	13	61	136	2
4	26	126	1	14	25	146	2
5	53	134	1	15	55	136	2
6	31	129	1	16	38	135	2
7	58	136	1	17	48	138	2
8	46	132	1	18	56	143	2

```

9      59      140      1      19      62      140      2
10     70      144      1      20      65      144      2
;
/*Frequentist Regression*/
proc reg data=one;
model sbp=age/clb;run;
ods graphics on;
proc mcmc data=one outpost=postout nbi=1000 nmc=10000 seed=123
plots=trace DIAGNOSTICS=ALL;
  *ods select Parameters REparameters PostSumInt tracepanel;
  parms b0 0 b1 0 s2 1;
  prior b: ~ normal(0, var = 10000);
  prior s: ~ igamma(0.01, scale = 0.01);
  mu = b0 + b1 * age;
  model SBP ~ normal(mu, var = s2);
  preddist outpred=pout nsim=1000;
run;
ods graphics off;

```

Table 9: SAS codes for simple regression with PROC MCMC

Bayesian MCMC output					
Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
b0	50000	119.6	4.3739	110.7	128.2
b1	50000	0.3219	0.0886	0.1533	0.5084
s2	50000	39.0260	14.6110	16.6404	67.4271

Output 4: The MCMC output regression of SBP on age.

Frequentist output					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	486.19921	486.19921	13.95	0.0015
Error	18	627.55079	34.86393		
Corrected Total	19	1113.75000			
Root MSE		5.90457	R-Square	0.4365	
Dependent Mean		134.75000	Adj R-Sq	0.4052	



Coeff Var	4.38187						
Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	95% Confidence Limits	
Intercept	1	119.88622	4.19352	28.59	<.0001	111.07596	128.69647
age	1	0.31659	0.08478	3.73	0.0015	0.13848	0.49470

**Output 5: The Frequentist output for regression of SBP on age.**

As we see the results from these approaches are pretty much the same as expected since we are using non-informative prior. Also for the prediction we can use the following codes

```
predist outpred=pout nsim=1000;
```

Portion of the output is given in the following table:

Posterior Summaries and Intervals for Prediction					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
SBP_1	1000	126.2	6.7909	113.4	140.1
SBP_2	1000	125.6	6.9519	112.6	139.4
SBP_3	1000	133.6	6.3207	120.9	145.5

We can use PROC GENMOD to do Bayesian regression as well as stated in the following codes.

```
proc genmod data=one;
model sbp=age;
bayes nbi=1000 nmc=10000 thin=10 seed=140;run;
```

Remark: in the case of random effect, we can easily extend the regression model as follows:

$$y_{ij} = \beta_0 + \beta_1 X_{ij} + \delta_i + \varepsilon_{ij}; \varepsilon_{ij} \sim N(0, \sigma^2) \text{ and } \delta_i \sim N(0, \tau^2)$$

is the random effect part of the model.  $i = 1, 2, \dots, k$  for  $i$ th cluster and  $j = 1, 2, \dots, n_i$  is number of observations in the  $i$ th group. The  $\beta_0$  and  $\beta_1$  are the fixed effects in regression coefficients.

Only one line of code will be added as follows:

```
random gamma ~ normal(0, var = s2g) subject=clin monitor=(gamma);
```

**Example8:** Nonlinear Poisson Regression Models: Taken from SAS/USER13.1.

This is an example of when PROC MCMC faces challenges to converge. This happens by mixing different proposal distributions. Also we need to do some transformations on some parameters. The data for this example come from SAS MCMC examples. It shows how to analyze count data for calls to a technical support help line in the weeks immediately following a product release. This information could be used to decide upon the allocation of technical support resources for new products. You can model the number of daily calls as a Poisson random variable, with the average number of calls modeled as a nonlinear function of **the number of weeks that have elapsed since the product's release.**

During the first several weeks after a new product is released, the number of questions that technical support receives concerning the product increases in a sigmoidal fashion. The expression for the mean value in the classic Poisson regression involves the log link. This an example shows the strength of PROC MCMC methodologies over the convenient Poisson models.

The mean function  $\lambda(t)$  is modeled as follows:

$$\lambda_i = \frac{\gamma}{1 + \exp[-(\alpha + \beta t_i)]}$$

Likelihood: 
$$L = \prod_{i=1}^n \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!}$$

Prior: Past experience with technical support data for similar products suggests the following prior distributions:

$$\gamma \sim \text{Gam}(3.5, 12) \quad \alpha \sim N(-5, 0.5^2) \quad \beta \sim N(0.75, 0.5^2)$$

**Posterior**  $\propto$  **likelihood**  $\times$  **Prior**

$$\begin{aligned} p(\gamma, \alpha, \beta | \text{data}) &\propto \prod_{i=1}^n \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \times \frac{(3.5)^{12}}{\Gamma(12)} \gamma^{12-1} e^{-3.5\gamma} \\ &\times \frac{1}{0.5\sqrt{2\pi}} \exp\left\{-\frac{1}{2 \times 0.5^2}\right\} (\alpha + 5)^2 \\ &\times \frac{1}{0.5\sqrt{2\pi}} \exp\left\{-\frac{1}{2 \times 0.5^2}\right\} (\beta - 0.75)^2 \end{aligned}$$

Nonlinear Poisson Model											
<code>data</code>	<code>calls;</code>										
<code>input</code>	<code>weeks calls @@;</code>										
<code>datalines;</code>											
1	0	1	2	2	2	2	1	3	1	3	3
4	5	4	8	5	5	5	9	6	17	6	9
7	24	7	16	8	23	8	27				

```

;
ods graphics on;
Title " Non-Linear Poisson model with Random effect statement";
proc mcmc data=calls outpost=callout seed=53197 ntu=1000
nmc=20000
propcov=quanew stats=none diag=ess;
ods select TADpanel ess;
parms alpha -4 beta 1 gamma 2;
prior gamma ~ gamma(3.5, scale=12);
prior alpha ~ normal(-5, sd=0.25);
prior beta ~ normal(0.75, sd=0.5);
lambda = gamma*logistic(alpha+beta*weeks);
model calls ~ poisson(lambda);run;
ods graphics off;

```

Table 10: SAS codes for Non-Linear Poisson regression via PROC MCMC

The diagnostic graphs, Figure 4, blow show clear issues with convergence and effective sample sizes (ESS) calculation in each parameters are all relatively low.

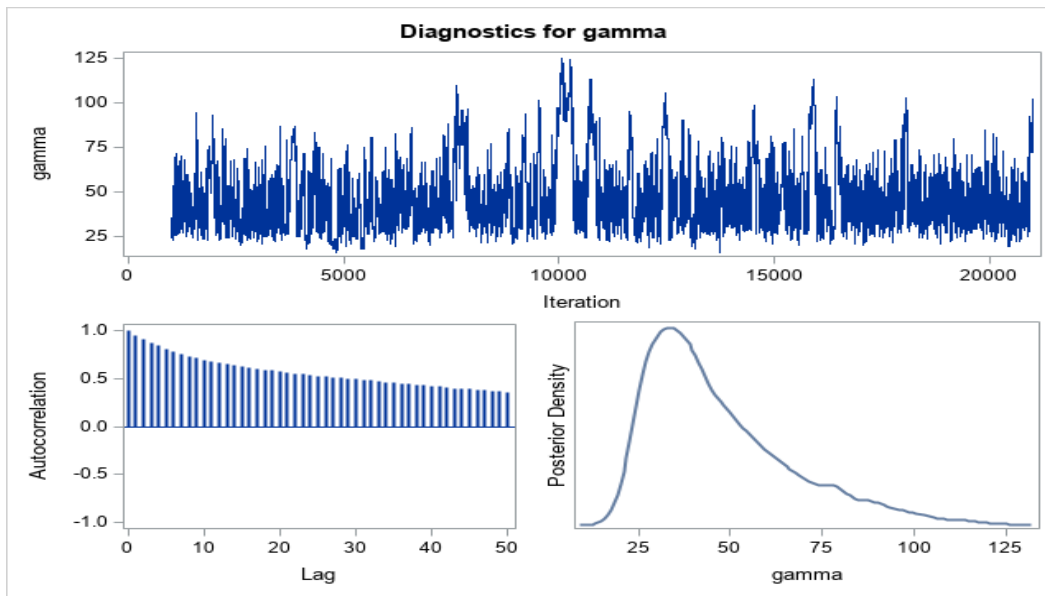


Figure 4 Diagnostic Plots for  $\gamma$

This is because a random walk Metropolis with a normal proposal is not always working well to get the joint distribution efficiently—the algorithm works best when the target distribution is unimodal and symmetric (normal-like). The nonlinearity in the parameters make it impossible to find a single proposal scale parameter that optimally adapts to different regions of the joint parameter space.

The parameter gamma has a positive support and the posterior distribution is right-skewed. Thus, let us use the logarithm of the parameter gamma for a better rate of convergence.

Take  $\delta = \ln(\gamma)$  and substitute in the above mean as follows:

$$\lambda_i = \frac{\exp(\delta)}{1 + \exp[-(\alpha + \beta t_i)]}, \text{ with the following priors}$$

$$\pi(\gamma) = \frac{1}{\Gamma(a)b^a} \gamma^{a-1} \exp\left\{-\frac{\gamma}{b}\right\}$$

$$\pi(\delta) = \frac{1}{\Gamma(a)b^a} (\exp(\delta))^{a-1} \exp\left\{-\frac{\exp(\delta)}{b}\right\} \times \exp(\delta)$$

Where the last term on the right is the Jacobian term of the transformation. This is called ExpGamma distribution which is supported by PROC MCMC. So we replace the gamma prior in the above with the following codes:

```
prior delta ~ egamma(3.5, scale=12);
```

You can obtain the same inference by specifying an ExpGamma prior on delta and take an exponential transformation to get back to gamma:

```
gamma = exp(delta);
```

Nonlinear Poisson Model With modified prior	
<pre>proc mcmc data=calls outpost=tcallout seed=53197 ntu=1000 nmc=20000 propcov=quanew diag=ess plots=(trace) monitor=(alpha beta gamma); ods select PostSumInt ESS TRACEpanel; parms alpha -4 beta 1 delta 2; prior alpha ~ normal(-5, sd=0.25); prior beta ~ normal(0.75, sd=0.5); prior delta ~ egamma(3.5, scale=12); gamma = exp(delta); lambda = gamma*logistic(alpha+beta*weeks); model calls ~ poisson(lambda);run;</pre>	

Table 11: SAS codes for Nonlinear Poisson model via proc MCMC

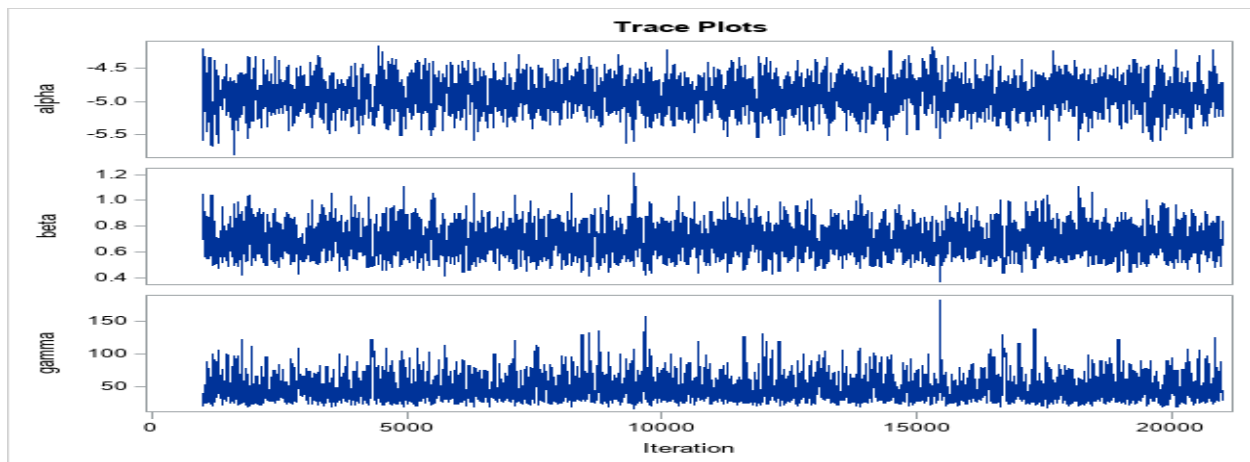


Figure 5: trace Plots for all 3 parameters

The trace plots in figure 5 show a better mixing of the parameters, and the effective sample sizes in the [output below](#) show substantial improvements over the original formulation of the model. The improvements are especially obvious in beta and gamma, where the increase is fivefold to tenfold.

**Example9 Meta-Analysis:** The meta-analysis by Yusuf et al. (1985) on the effectiveness of beta blockers for reducing mortality and reinfarction is usually cited as the reference for what is called modified Mantel-Haenszel method for meta-analyzing data from 22 such studies. At each center, patients were randomly assigned to receive beta blockers or a placebo after a myocardial infarction, in the form of  $2 \times 2$  tables and the mortality rates were recorded. The outcome measure is the estimated log-odds ratio under a random-effects model due to heterogeneity of studies. When we have no prior information for believing that any particular study is different from another (exchangeability), you can treat Bayesian meta-analysis as a hierarchical model Chen, F. (2011). The following variables read the data into SAS:

Center: Study id,  $i = 1, 2, \dots, 22$ .

$x_t$ : Number of deaths in treatment group       $n_t$ : Number of patients in treatment group

$x_c$ : Number of deaths in control group       $n_c$ : Number of patients in control group

Assumptions and Notations:

$x_t \sim \text{Bin}(n_t, p_{ti})$ , where  $p_{ti}$  is the probability of death in treatment group.

$x_c \sim \text{Bin}(n_c, p_{ci})$ , where  $p_{ci}$  is the probability of death in control group.

$\phi_i = \ln\left(\frac{p_{ci}}{1-p_{ci}}\right)$  is the log odds of control group.

$\phi_i + \theta_i = \ln\left(\frac{p_{ti}}{1-p_{ti}}\right)$  is the log odds of treatment group.

Then  $\theta_i = \ln\left(\frac{p_{ti}}{1-p_{ti}}\right) - \ln\left(\frac{p_{ci}}{1-p_{ci}}\right) = \ln\left(\left(\frac{p_{ti}}{1-p_{ti}}\right) / \left(\frac{p_{ci}}{1-p_{ci}}\right)\right)$  is the log of odds ratio or

treatment effect. Furthermore assume  $\phi_i \sim N(\mu_\phi, \sigma_\phi^2)$  and  $\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$

With the following priors

$\mu_\phi, \mu_\theta \sim N(0, 3)$  and  $\frac{1}{\sigma_\phi^2}, \frac{1}{\sigma_\theta^2} \sim \text{Gamma}(1, 2)$

Center	$x_t$	$n_t$	$x_c$	$n_c$	OR
1	3	38	3	39	1.02
2	7	114	14	116	0.54

3	5	69	11	93	0.64
4	102	1533	127	1520	0.81
5	28	355	27	365	1.06
6	4	59	6	52	0.61
7	98	945	152	939	0.67
8	60	632	48	471	0.94
9	25	278	37	282	0.71
10	138	1916	188	1921	0.75
11	64	873	52	583	0.83
12	45	263	47	266	0.97
13	9	291	16	293	0.58
14	57	858	45	883	1.28
15	25	154	31	147	0.80
16	33	207	38	213	0.91
17	28	251	12	122	1.12
18	8	151	6	154	1.34
19	6	174	3	134	1.52
20	32	209	40	218	0.86
21	27	391	43	364	0.61
22	22	680	39	674	0.57
	826	10441	985	9849	0.82

**Table 12: Data from 22 different studies to instigates Odds Ratio**

An easy frequentist estimate of meta-analysis of this kind of data is given by

$$OR_{pooled} = \frac{\sum_{i=1}^{22} \left( \frac{x_{t_i} n_{c_i}}{n_{t_i} + n_{c_i}} \right)}{\sum_{i=1}^{22} \left( \frac{x_{c_i} n_{t_i}}{n_{t_i} + n_{c_i}} \right)} = 0.8238.$$

This assumes homogenous data across all studies. However, in general this assumption may not hold.

Meta Analysis of Multi-Centers.
<pre> data meta; input center Xc Nc Xt Nt; datalines; </pre>

```

1 3 039 3 038
2 14 116 7 114
3 11 093 5 069
..
;
proc mcmc data=meta outpost=PostOut nmc=50000 thin=5 monitor=(OR Pooled);
parms mu_theta mu_phi s_theta s_phi;
prior mu: ~ normal(0,sd=3);
prior s: ~ igamma(0.01,s=0.01);
random theta ~normal(mu_theta, var=s_theta) subject=center; /*1*/
random phi ~normal(mu_phi, var=s_phi) subject=center;
pc = logistic(phi);
model Xc ~ binomial(Nc, pc);
pt = logistic(theta + phi);
model Xt ~ binomial(Nt, pt);
array OR[22];
OR[center]=exp(theta);
Pooled=exp(mu_theta);
run;

```

Table 13: SAS codes for meta-analysis via PROC MCMC, SAS support.

The first RANDOM statements is  $\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$  and the second one  $\phi_i \sim N(\mu_\phi, \sigma_\phi^2)$ . The SUBJECT= center is the grouping index for the centers. The statement, OR [center]=exp(theta), calculates the odds ratio for each center as it is shown in the output below as a posterior summary. The symbol Pooled calculates the overall odds ratio for the treatment effect.

Table 14 shows the effect of treatment varies from study to study, the most effective is at center 7 (OR= 0.6889) and least effective is at center 14 (OR= 0.9340). The pooled posterior estimate of the odds ratio is 0.7824 (95% HPD CI= 0.6842, 0.8838).

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
OR1	10000	0.7985	0.1390	0.5361	1.0795
OR2	10000	0.7514	0.1216	0.5065	0.9850
OR3	10000	0.7748	0.1269	0.5332	1.0403
OR4	10000	0.7817	0.0805	0.6237	0.9389
OR5	10000	0.8428	0.1290	0.6119	1.1098
OR6	10000	0.7782	0.1304	0.5225	1.0479
OR7	10000	<b>0.6889</b>	<b>0.0761</b>	<b>0.5334</b>	<b>0.8305</b>
OR8	10000	0.8371	0.1124	0.6465	1.0812
OR9	10000	0.7603	0.1086	0.5530	0.9833
OR10	10000	0.7434	0.0706	0.6040	0.8803
OR11	10000	0.7963	0.0986	0.6128	0.9993
OR12	10000	0.8566	0.1249	0.6311	1.1084
OR13	10000	0.7375	0.1177	0.5073	0.9719
OR14	10000	<b>0.9340</b>	<b>0.1436</b>	<b>0.6868</b>	<b>1.2297</b>
OR15	10000	0.7977	0.1194	0.5703	1.0404
OR16	10000	0.8293	0.1212	0.6094	1.0742
OR17	10000	0.8502	0.1387	0.5957	1.1285
OR18	10000	0.8078	0.1365	0.5613	1.0937
OR19	10000	0.7865	0.1294	0.5453	1.0578

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
OR20	10000	0.8125	0.1170	0.5956	1.0513
OR21	10000	0.7229	0.1039	0.5159	0.9312
OR22	10000	0.7072	0.1029	0.5051	0.9118
<b>Pooled</b>	10000	<b>0.7824</b>	<b>0.0516</b>	<b>0.6842</b>	<b>0.8838</b>

Table 14: The OR for each study and for the pooled data

## CONCLUSION

The Bayesian approach via PROC MCMC presented in this paper is for the introduction of any model building approach. We are trying to show to the readers how easy it is to use SAS to implement Bayesian models without that much of code writing. We started with very basic and elementary examples to make sure you will understand how MCMC works and how easily you can manipulate the SAS code to do some difficult analysis with some minor modifications. PROC GENMOD is an example, however, as a warning to people who have no foundation in Bayesian analysis. We are hoping after you went through all the examples you will agree that SAS is an easy tool for Bayesian analysis, especially for teaching students who have difficulties with programming. Although we will recommend that for real research one must be extra careful in the use of these procedures, in particular the choice of priors and the transformation of the parameters. Example 11 in this paper is an evidence that how dangerous it can be if you are unfamiliar with MCMC. It was also shown that the Bayesian approach overcomes difficulties of convergence common in the frequentist approach for the many models like log-binomial model. For more details please see the article by Bielefeldt et al. (2015) for more details. It correctly restricts (ie, choose appropriate prior) the parameter space to produce valid probabilities.

## REFERENCES

- Bielefeldt V, Torman L, and Suzi Alves Camey SA (2015). Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes, *Themes in Epidemiology*.
- Casella, G. and George, E. (1992) "Explaining the Gibbs Sampler," *Amer. Statistician* 46, 167-174.
- Casella, G., & Berger, R. L. (2002). *Statistical inference*. Belmont, CA: Duxbury.**
- Chen, F. (2009), "Bayesian Modeling Using the MCMC Procedure," in *Proceedings of the SAS Global Forum 2009 Conference*, Cary, NC: SAS Institute Inc.



Chen, F. (2011). "The RANDOM Statement and More: Moving On with PROC MCMC." In Proceedings of the SAS Global Forum 2011 Conference. Cary, NC: SAS Institute Inc.

**Gelfand, A. E., and A. F. M. Smith (1989). "Sampling-Based Approaches to Calculating Marginal Densities", Jour. Amer. Stat. Assn. 85,398-409.**

Gelfand, A. E., Hills, S. E., Racine-Poon. A., and Smith, A. F. M. (1990) . Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of American Statistical Association* 85. 972-985.

**Geman, A. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.**

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bayesian Statistics 4, J. M. Bernardo, J. O. Berger, A P. Dawid, and A. F. M. Smith (eds), 169-193. Oxford: Clarendon Press.

Hastings W.K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications, "*Biometrika*, 57, 97-109.

Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York:Wiley; 2000.

McNutt LA, Wu C, Xue X, Hafner JP (2003). Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. *Am J Epidemiol*; 157(10):940-3. 2. 4

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., and Teller E. (1953), "Equations of

Pedroza and Truong (2016) Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes, *BMC Medical Research Methodology* (2016).

Pedroza and Truong (2017), Estimating relative risks in multicenter studies with a small number of centers which methods to use? A simulation study, *Trials*.

**SAS Institute Inc. 2013. SAS/STAT® 13.1 User's Guide.**

SAS SUPPORT: BAYESIAN HIERARCHICAL MODELING FOR META-ANALYSIS

[https://support.sas.com/rnd/app/stat/examples/BayesMeta/new\\_example/index.html](https://support.sas.com/rnd/app/stat/examples/BayesMeta/new_example/index.html).

**Tanner, M. A., and Wong. W.(1987). "The Calculation of Posterior Distribution by Data Augmentation "( with discussion), *Journal of American Statistical Association* 85. 528-550.**

UCLA: <https://stats.idre.ucla.edu/sas/faq/how-can-i-estimate-relative-risk-in-sas-using-proc-genmod-for-common-outcomes-in-cohort-studies/>.

Yusuf, S., Peto, R., Lewis, J., Collins, R., & Sleight, P. (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Disease*, 27(5), 335–371.

Zeger, S.L. and Karim, M.R. (1991). Generalized Linear Models with random effects: A Gibbs sampling approach. Journal of the American Statistical Association 86,79-86.

#### CONTACT INFORMATION

In the event of any questions, comments, or whatever, you can contact the first author via email:

Mark Ghamsary  
Global Statistical Consulting Group  
Email: [mghamsary@gmail.com](mailto:mghamsary@gmail.com)

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 trademarks of their respective companies.