Bayesian Modeling in SAS Software

June 19, 2007

BMI data

Lewis and Taylor (1967) conducted a cross-sectional study of the age (in years), height (in inches), and weight (in pounds) of $n = 237$ students (ranging in age from 13.9 to 25 years of age, with a median of 16.3 years). Body mass index, calculated as weight (in kg) divided by the square of height (in cm), is often used to identify patients who are at risk for overweight or underweight status.

BMI data

Fashion week in Madrid, Spain recently used a cutoff of BMI=18 as a marker that runway fashion models were underweight (these women were not allowed to model and had dietary counseling made available to them). Around 30% of models were turned away due to not meeting this requirement. We will apply this cutoff in our analysis of underweight status of adolescents. (For comparison, my BMI is 20.5, and I would need to lose 15 pounds for my BMI to be less than 18.)

Weight data

We consider a model for the indicator of at risk for underweight status, z_i, as a function of age and gender.

We will consider the model

$$logit\left(Pr(z_i = 1)\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i},$$

in which x_1 represents age (centered at age 16), and x_2 takes value 1 for males and 0 for females. Here,

$$\pi_i = \frac{\exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right)}{1 + \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right)}.$$

Weight data

For the uncentered weight data, we have

$$\mathbf{z} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}_{237\times 1}, \quad \mathbf{X} = \begin{pmatrix} 1 & 14.3 & 0 \\ 1 & 15.5 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 15.1 & 1 \end{pmatrix}_{237\times 3},$$

and $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$. Before analysis, we will center age by subtracting 16 from every subject's age value. This should help with convergence.

Frequentist analysis of weight data

We start by carrying out a standard frequentist analysis.

```
proc logistic data=htwt3 descending;
  model underweight=age_c male;
run;
```

Frequentist analysis of weight data

Parameter	*Estimate*	*95% CI*
$\exp(\beta_1)$	0.80	(0.68, 0.92)
$\exp(\beta_2)$	1.08	(0.64, 1.82)

We see no evidence of a relationship between gender and BMI, though a one-year increase in age is associated with 0.80 the odds of being underweight (so that younger adolescents are skinnier).

Prior distribution

We use our investigator's knowledge of the subject matter in specifying a normal prior distribution. Consider the intercept, β_0, which represents the logit of the probability that a sixteen-year old girl has BMI<18. Assuming $\beta_0 \sim N(0, 1)$ implies that we are pretty sure that the probability of underweight status (BMI<18) for this subject is not less than

$$100 \times \frac{\exp(0 - 2 \times 1)}{1 + \exp(0 - 2 \times 1)} = 12\%$$

or greater than

$$100 \times \frac{\exp(0 + 2 \times 1)}{1 + \exp(0 + 2 \times 1)} = 88\%.$$

(For the N(0,1) distribution, 95.4% of the probability lies between 2 and -2.)

Prior distribution

For the log odds ratios β_1 and β_2, we use independent $N(0, 0.35)$ prior distributions to express the beliefs of investigators that the odds ratios less than

$$\exp\left(0 - 2 \times \sqrt{0.35}\right) = 0.31$$

or greater than

$$\exp\left(0 + 2 \times \sqrt{0.35}\right) = 3.26$$

are unanticipated.

Prior distribution

Because the mean determines the variance for the binomial distribution, we do not have a scale parameter ϕ for which we must specify a prior. (Here, $\phi = 1$.)

For other distributions, notably the normal distribution (linear regression), one must specify a prior distribution for a scale parameter, or a function of it (like the variance σ^2 in a normal linear regression model – the scale parameter here is just σ, and often we place the prior on the precision, which is $\frac{1}{\sigma^2}$).

SAS software

SAS STAT software now provides Bayesian analysis in downloadable, experimental versions of three procedures for SAS 9.1.3 on Windows: GENMOD, LIFEREG, and PHREG. The new BAYES statement in these procedures produces Bayesian modeling and inference capability in generalized linear models, accelerated life failure models, Cox regression models, and piecewise constant baseline hazard models (also known as piecewise exponential models). These versions are named BGENMOD, BLIFEREG, and BPHREG, respectively, and they otherwise contain the full functionality of the original procedures.

SAS software

The developers and user support are outstanding, with more exciting developments in the works.

Visit the URL below for documentation and to download the experimental versions of the procedures.

`www.sas.com/statistics`

Specifying prior in SAS

```
data NormalPrior;
  input _type_ $ Intercept age_c male;
  datalines;
  Var 1 0.35 0.35
  Mean 0 0.0 0.0
  ;
run;
```

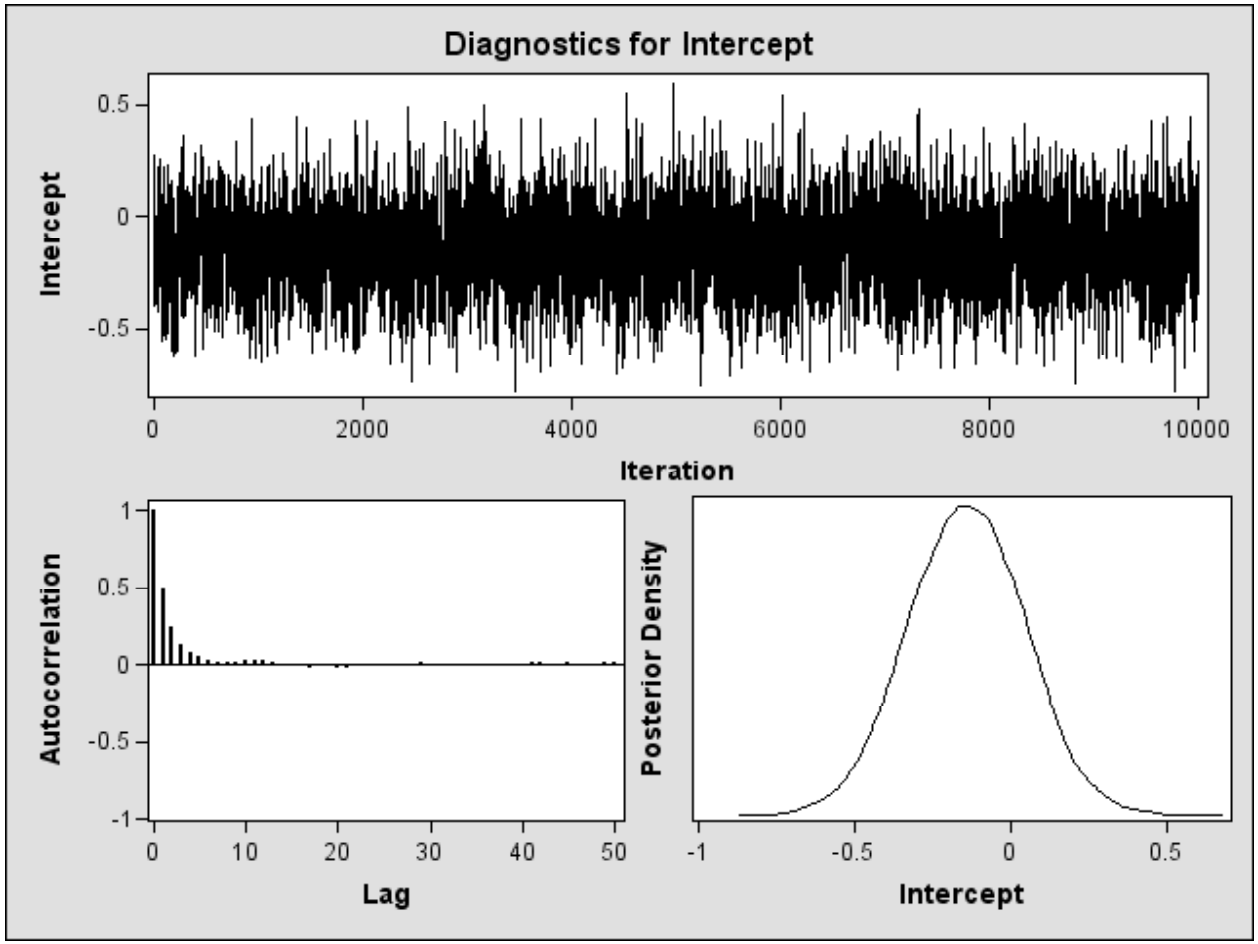
Specifying Bayesian analysis in SAS

```
proc bgenmod descending data=htwt3;
  model underweight=age_c male/dist=binomial link=logit;
  bayes seed=2345 plots=all nbi=2000 nmc=10000
     coeffprior=normal(input=NormalPrior);
  ods output PosteriorSample=Post;
run;
```

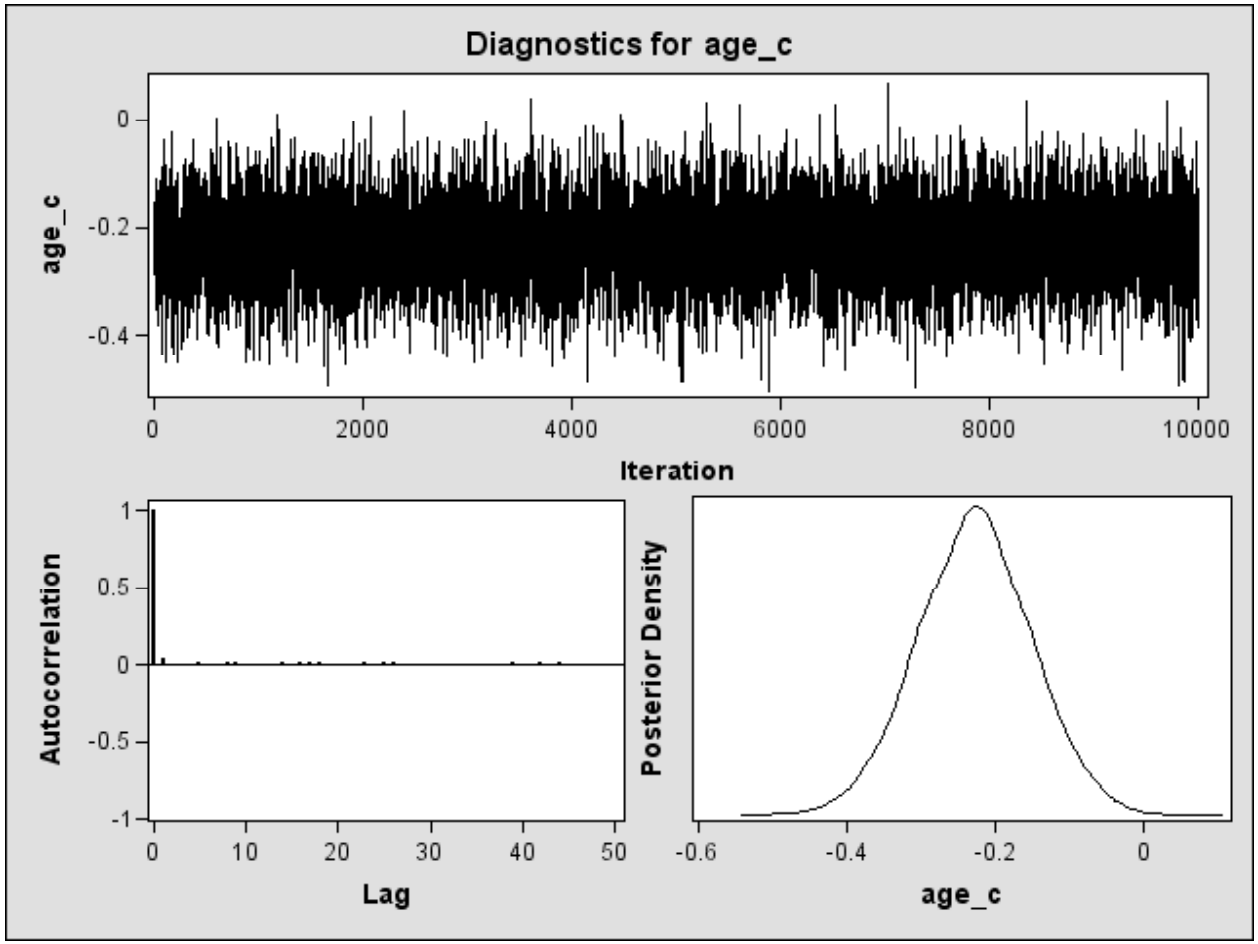
SAS output

Next we'll walk through all the SAS output in the order in which it is printed. We'll start by checking convergence visually via the graphics produced by PROC BGENMOD.

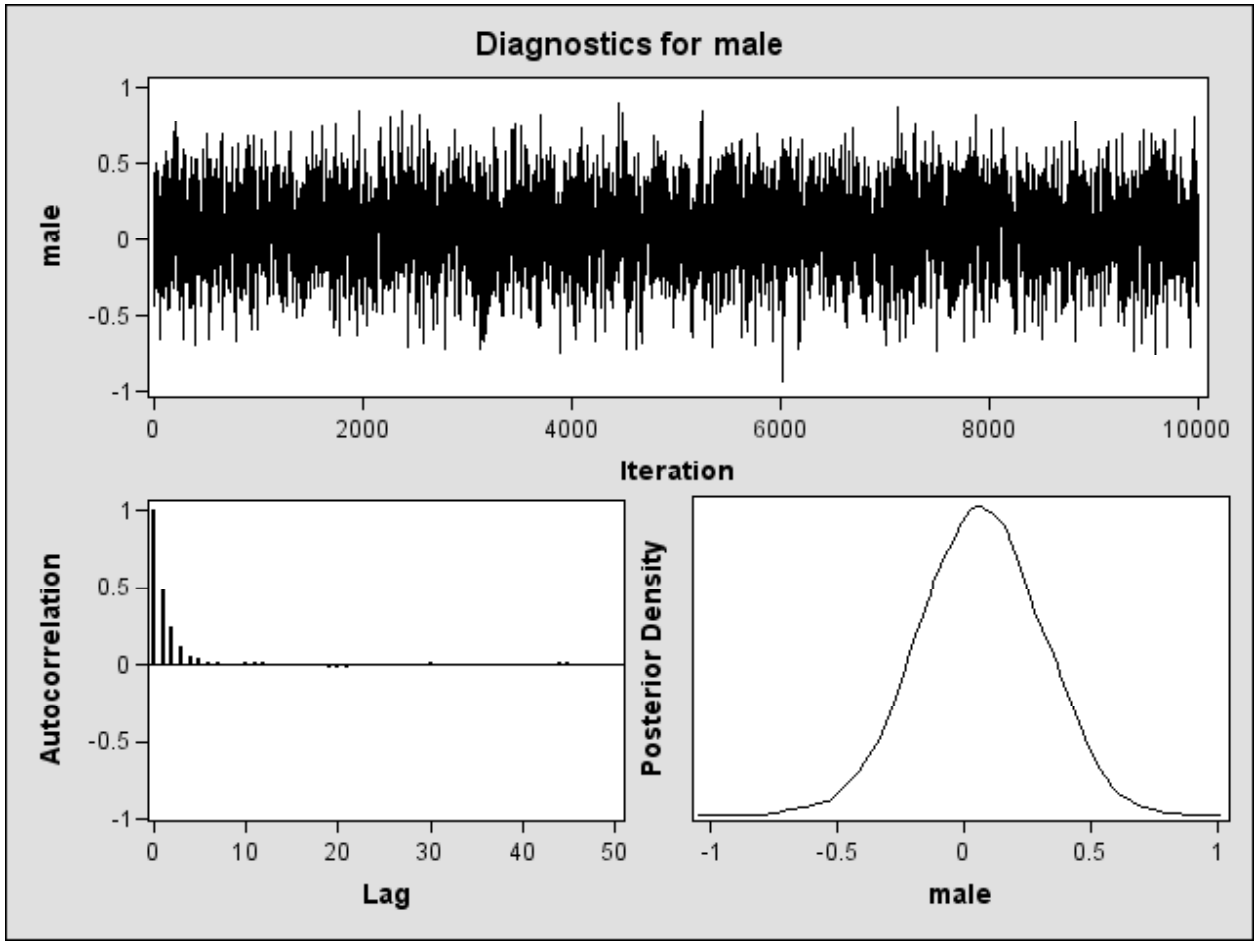
Graphical assessments for intercept



Graphical assessments for age



Graphical assessments for gender



SAS output

Next we'll move to the output appearing in our *.lst* file.

SAS output

The BGENMOD Procedure
Bayesian Analysis

Model Information

Data Set	WORK.HTWT3
Burn-In Size	2000
MC Sample Size	10000
Thinning	1
Distribution	Binomial
Link Function	Logit
Dependent Variable	underweight

Number of Observations Read	237
Number of Observations Used	237
Number of Events	107
Number of Trials	237

SAS output

SAS begins by giving us some basic information about the model. Information unique to the Bayesian procedures includes the burn-in size. For this analysis, we specified the default burn-in of 2000 samples, though we can increase this if necessary or desired by using the *nbi=* option on the **BAYES** statement line. Similarly, we used 10000 MCMC samples for posterior inference after the burn-in (for a total of 12000 samples), and we can increase this number using the *nmc=* option. We did not thin the chain (thinning=1); thinning is useful if we are not able to store a large number of samples and need to improve convergence. To thin the chain, use the option *THINNING=k* on the **BAYES** statement line to keep every kth sample.

SAS output

Response Profile

Ordered Value	underweight	Total Frequency
1	1	107
2	0	130

PROC BGENMOD is modeling the probability that underweight='1'.

SAS output

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	234	316.6524	1.3532
Scaled Deviance	234	316.6524	1.3532
Pearson Chi-Square	234	235.5328	1.0066
Scaled Pearson X2	234	235.5328	1.0066
Log Likelihood		-158.3262	
Full Log Likelihood		-158.3262	

Algorithm converged.

SAS output

The previous output is based on the frequentist analysis.

SAS output

The BGENMOD Procedure

Bayesian Analysis

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept	1	-0.1459	0.1967	-0.5314	0.2397
age_c	1	-0.2291	0.0766	-0.3792	-0.0791
male	1	0.0767	0.2670	-0.4466	0.6001
Scale	0	1.0000	0.0000	1.0000	1.0000

NOTE: The scale parameter was held fixed.

SAS output

First, we get the frequentist analysis by default.

SAS output

Bayesian Analysis
Independent Normal Prior for Regression Coefficients

Parameter	Mean	Precision
Intercept	0	1
age_c	0	2.857143
male	0	2.857143

Initial Values and Seeds

Chain	_SEED_	Intercept	age_c	male
1	2345	-0.14479	-0.22902	0.04228

SAS output

Here, we see the prior we used. We assumed prior independence of β_0, β_1, and β_2, with prior means 0 for all parameters and prior variance 1 for β_0 and prior variance $0.35 = \frac{1}{2.857143}$ for β_1 and β_2. SAS also provides us with the seed (so that we can reproduce the analysis later), the number of chains used, and the starting values for the parameters. Use the *SEED=* option on the **BAYES** statement line to change the seed. While you can use multiple chains in order to calculate the Gelman-Rubin convergence diagnostic, you cannot get inferences based on multiple chains (without programming this yourself by using different seeds and combining across multiple runs of the procedure...wouldn't be too difficult if you were so inclined).

SAS output

Bayesian Analysis

Descriptive Statistics of the Posterior Samples

Param	N	Mean	Standard Deviation	Quantiles 25%	50%	75%
Int.	10000	-0.1398	0.1865	-0.2680	-0.1395	-0.0136
age_c	10000	-0.2286	0.0764	-0.2803	-0.2278	-0.1775
male	10000	0.0638	0.2427	-0.0980	0.0642	0.2269

SAS output

Interval Statistics of the Posterior Samples

Parameter	Alpha	Credible Interval		HPD Interval	
Intercept	0.050	-0.5064	0.2268	-0.5043	0.2269
age_c	0.050	-0.3792	-0.0780	-0.3818	-0.0819
male	0.050	-0.4230	0.5279	-0.4045	0.5402

SAS output

SAS provides the mean, standard deviation, and quartiles of the sampled values for each parameter. In addition, SAS outputs a 95% credible interval and 95% HPD interval for each parameter. To change the coverage of the intervals, you can use the *SUMMARIES* option on the **BAYES** statement. For example, *SUMMARIES=(DESCRIPTIVE INTERVAL(ALPHA=0.10))* would give you 90% intervals.

SAS output

Autocorrelations of posterior samples as convergence diagnostics here are not worrisome; they are all 0.05 or less by lag 5.

The BGENMOD Procedure

Bayesian Analysis

Autocorrelations of the Posterior Samples

Parameter	Lag1	Lag5	Lag10	Lag50
Intercept	0.4895	0.0501	0.0311	0.0146
age_c	0.0419	0.0192	-0.0038	0.0052
male	0.4756	0.0360	0.0096	0.0016

SAS output

Geweke Diagnostics

Parameter	z	Pr > |z|
Intercept	0.0669	0.9467
age_c	-0.2751	0.7832
male	-0.2754	0.7830

The Geweke test (comparing values in the early part of the Markov chain to those in the latter part) statistics are all fairly close to zero, providing no evidence of lack of convergence.

SAS output

Effective Sample Size

Parameter	ESS	Correlation Time	Efficiency
Intercept	3259.1	3.0684	0.3259
age_c	9226.8	1.0838	0.9227
male	3476.9	2.8761	0.3477

The effective sample size and autocorrelation time are related to mixing of the Markov chain. Large discrepancies between the effective sample size and the actual number of samples taken (here, 10000) can be indicative of poor mixing.

Posterior probabilities and CI's for OR

Posterior probabilities and credible intervals for functions of the parameter estimates can be obtained easily using the posterior samples (the code output them to the file *Post*). In particular, we are interested in the posterior probabilities that age and gender are protective for underweight status. In addition, we would like to calculate 95% credible intervals for the odds ratios.

SAS code for posterior probabilities

```
data postprocess; set Post;
  ageind=(age_c<0);
  maleind=(male<0);
  ageor=exp(age_c);
  maleor=exp(male);
run;
/* get posterior probabilities of interest */
proc univariate; var ageind maleind ageor maleor; run;
/* get credible interval for odds ratio */
proc univariate data=postprocess;
  var ageor maleor;
  output out=Pctls pctlpts=2.5 97.5
     pctlpre=ageor_ maleor_ pctlname=LCL UCL;
run;
proc print data=Pctls; run;
```

Bayesian analysis of weight data

Parameter	*Post. mean*	*95% CI*	$Pr(\exp(\beta) < 1)$
$\exp(\beta_1)$	0.80	(0.68, 0.93)	0.9987
$\exp(\beta_2)$	1.10	(0.66, 1.70)	0.3907

We see no evidence of a relationship between gender and BMI, though a one-year increase in age is associated with 0.80 the odds of being underweight (so that younger adolescents are skinnier).

Impact of prior on posterior distribution

- Many of you may be worried about prior specification and impact of the prior on the posterior distribution.

Impact of prior on posterior distribution

- Many of you may be worried about prior specification and impact of the prior on the posterior distribution.
- To illustrate impact of the prior on the posterior, we will concentrate on the coefficient for age, β_1.

Impact of prior on posterior distribution

- Many of you may be worried about prior specification and impact of the prior on the posterior distribution.
- To illustrate impact of the prior on the posterior, we will concentrate on the coefficient for age, β_1.
- Frequentist result: $\exp\left(\widehat{\beta}_1\right) = 0.80$ with 95% CI=$(0.68, 0.92)$.

Impact of prior on posterior distribution

- Many of you may be worried about prior specification and impact of the prior on the posterior distribution.
- To illustrate impact of the prior on the posterior, we will concentrate on the coefficient for age, β_1.
- Frequentist result: $\exp\left(\widehat{\beta}_1\right) = 0.80$ with 95% CI=$(0.68, 0.92)$.
- We will consider models using the normal prior specification $\pi(\boldsymbol{\beta}) = N\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0\right)$.

Reference analysis

As a reference point, we will use the results from the "non-informative" proper prior

$$\pi(\boldsymbol{\beta}) = N\left(\boldsymbol{\beta}; \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, 10^6 \mathbf{I}_3\right),$$

which is the default normal prior in SAS.

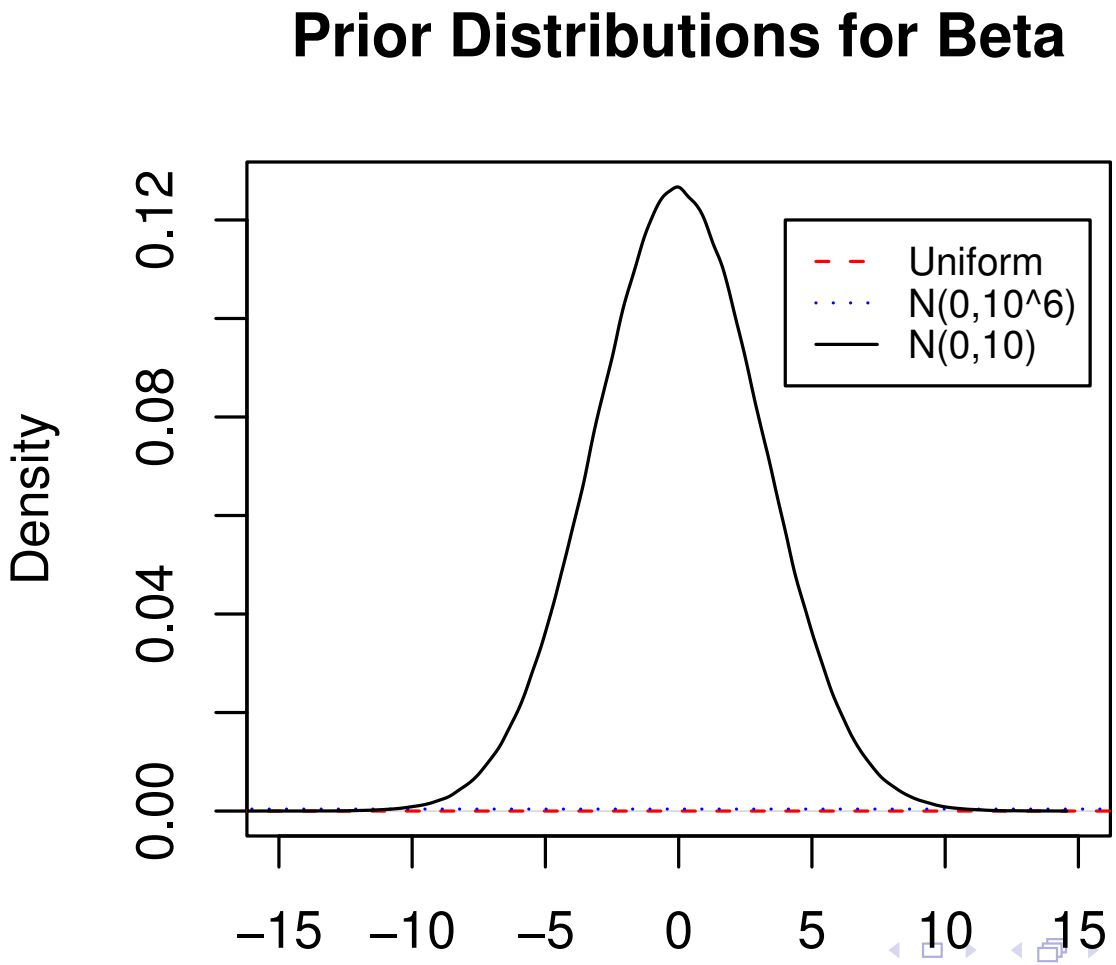
Previously, we used the prior

$$N\left(\boldsymbol{\beta}; \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.35 & 0 \\ 0 & 0 & 0.35 \end{pmatrix}\right).$$

"Noninformative" priors

Many researchers use the term "non-informative" for priors that are very spread out. However, in many ways, this type of prior is not necessarily a good choice. The following figure shows two "non-informative" priors, a normal prior with a big variance and a uniform improper prior, for β as well as a more informative prior (normal prior with smaller variance).

"Noninformative" priors



"Noninformative" priors

As you can see, the "noninformative" priors imply something unrealistic in most epidemiologic applications. In particular, the normal prior with variance 10^6 implies that the prior probability of an odds ratio between 0.14 and 7.39 (log OR between -2 and 2) is only 0.0016, while the normal prior with variance 10 implies that this prior probability is 0.47. Thus in many epidemiologic applications, more informative priors simply make more sense. (In fact, you might want to make this one tighter.)

Misspecified prior mean

Suppose we misspecify the prior mean for $\boldsymbol{\beta}$ so that we use

$$\pi(\boldsymbol{\beta}) = N\left(\begin{pmatrix} 0 \\ 0.69 \\ 0 \end{pmatrix}, \begin{pmatrix} 1000000 & 0 & 0 \\ 0 & 1000000 & 0 \\ 0 & 0 & 1000000 \end{pmatrix}\right).$$

Inference on $\exp(\beta_1)$

Prior	*Posterior Mean*	*Posterior Median*	*95% Credible Interval*
$N(0, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 10^6)$	0.79	0.79	(0.68, 0.92)

Because our prior variance is huge, the mis-specified mean is not a problem.

Misspecified prior mean

Suppose we misspecify the prior mean and variance for $\boldsymbol{\beta}$ and use

$$\pi(\boldsymbol{\beta}) = N\left(\begin{pmatrix} 0 \\ 0.69 \\ 0 \end{pmatrix}, \begin{pmatrix} 1000000 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1000000 \end{pmatrix}\right).$$

This implies that odds ratios less than $\exp(0.69 - 2) = 0.27$ and greater than $\exp(0.69 + 2) = 14.73)$ are unanticipated.

Inference on $\exp(\beta_1)$

Prior	*Posterior Mean*	*Posterior Median*	*95% Credible Interval*
$N(0, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 1)$	0.80	0.80	(0.68, 0.92)

Because our prior variance is still relatively large, the mis-specified mean is not yet a problem.

Misspecified prior mean

Suppose we misspecify the prior mean and variance for $\boldsymbol{\beta}$ and use

$$\pi(\boldsymbol{\beta}) = N\left(\begin{pmatrix} 0 \\ 0.69 \\ 0 \end{pmatrix}, \begin{pmatrix} 1000000 & 0 & 0 \\ 0 & 0.16 & 0 \\ 0 & 0 & 1000000 \end{pmatrix}\right).$$

This now implies that odds ratios less than $\exp(0.69 - 2(0.4)) = 0.90$ and greater than $\exp(0.69 + 2(0.4)) = 4.44$ are unanticipated.

Inference on $\exp(\beta_1)$

Prior	*Posterior Mean*	*Posterior Median*	*95% Credible Interval*
$N(0, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 1)$	0.80	0.80	(0.68, 0.92)
$N(0.69, 0.16)$	0.82	0.81	(0.70, 0.95)

Because our prior variance is still relatively large relative to the amount of information in the data, the mis-specified mean is not yet a problem.

Misspecified prior mean

Suppose we misspecify the prior mean and variance for $\boldsymbol{\beta}$ and use

$$\pi(\boldsymbol{\beta}) = N\left(\begin{pmatrix} 0 \\ 0.69 \\ 0 \end{pmatrix}, \begin{pmatrix} 1000000 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 1000000 \end{pmatrix}\right).$$

This now implies that odds ratios less than $\exp(0.69 - 2(0.1)) = 1.63$ and greater than $\exp(0.69 + 2(0.1)) = 2.44$ are unanticipated. This is approaching a *point mass* prior (putting all our prior probability at 2).

Inference on $\exp(\beta_1)$

Prior	*Posterior Mean*	*Posterior Median*	*95% Credible Interval*
$N(0, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 10^6)$	0.79	0.79	(0.68, 0.92)
$N(0.69, 1)$	0.80	0.80	(0.68, 0.92)
$N(0.69, 0.16)$	0.82	0.81	(0.70, 0.95)
$N(0.69, 0.01)$	1.09	1.09	(0.97, 1.21)

Now we start seeing an effect on the posterior mean, but still the data are pulling us back in the right direction.

Programs and Data

Data and SAS programs for all analyses presented will be available at the NIEHS Biostatistics Branch website.

`http://dir.niehs.nih.gov/dirbb/serbayes/`