Bayesian Binomial Model with Power Prior Using the MCMC Procedure

Overview

Power priors (Ibrahim and Chen 2000) are a useful family of informative priors when relevant historical data is available. You can use power priors when you want to take historical data into account while analyzing similar, current data. This example illustrates fitting a Bayesian binomial model with a power prior and also demonstrates two useful features of PROC MCMC: using nonstandard distributions and assigning different likelihood functions to different subsets of the data.

The SAS source code for this example is available as an attachment in the text file. In Adobe Acrobat, right-click the icon in the margin and select **Save Embedded File to Disk**. You can also double-click to open the file immediately. source code

Analysis

Suppose that, in addition to data from a current study, you have similar data from a historical data set or pilot study. A power prior enables you to capture the information collected in the pilot study and incorporate it into analysis and inference in the current study. The historical data gets integrated via an informative prior and provides a convenient way of updating past knowledge.

A power prior can be expressed as a product of the weighted likelihood of $\boldsymbol{\theta}$ conditional on the historical data and a prior distribution on the parameters before any data is observed. More formally, the power prior distribution of $\boldsymbol{\theta}$ is written as

$$\pi(\boldsymbol{\theta}|D_0, a_0) \quad \propto \quad L(\boldsymbol{\theta}|D_0)^{a_0} \pi_0(\boldsymbol{\theta})$$

where $D_0 = (n_0, \mathbf{y}_0, \mathbf{X}_0)$ are the historical data (from the pilot study), $L(\boldsymbol{\theta}|D_0)$ is the likelihood of $\boldsymbol{\theta}$ conditional on historical data, a_0 is a discounting or scale parameter with $0 \leq a_0 \leq 1$ that controls the amount of weight you want to put on the historical data, and $\pi_0(\boldsymbol{\theta})$ is a prior for $\boldsymbol{\theta}$ before the historical data D_0 are observed.

The posterior distribution of $\boldsymbol{\theta}$ is

$$
\begin{aligned}
\pi(\boldsymbol{\theta}|D, D_0, a_0) \quad &\propto \quad L(\boldsymbol{\theta}|D)\pi(\boldsymbol{\theta}|D_0, a_0) \\
&\propto \quad L(\boldsymbol{\theta}|D)L(\boldsymbol{\theta}|D_0)^{a_0}\pi_0(\boldsymbol{\theta}) \\
&\propto \quad \prod_{i=1}^{n} f(y_i|\theta, x_i) \prod_{j=1}^{n_0} f(y_{0,j}|\theta, x_{0,j})^{a_0}\pi_0(\theta)
\end{aligned}
\tag{1}
$$

where $D = (n, \mathbf{y}, \mathbf{X})$ represents the data from the current study, $\mathbf{y} = \{y_i\}, \mathbf{x} = \{x_i\}$ for $i = 1 \cdots n$, $\mathbf{y_0} = \{y_{0,j}\}, \mathbf{x_0} = \{x_{0,j}\}$ for $j = 1 \cdots n_0$ represents the historical data, and $f(\cdot)$ represents the density function for a single observation in either the historical or the current data set.

You can combine the two data sets to form a new data set D^* and rewrite the posterior distribution in Equation 1 as the following:

$$
p(\theta|D^*, a_0) \quad \propto \quad \prod_{i=1}^{n+n_0} f_i(y_i|\theta, x_i) \cdot \pi_0(\theta)
\tag{2}
$$

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

The posterior distribution Equation 2 is used to model the hypothetical clinical trial setting described in this example. The following SAS statements create the PTRIALS data set that represents binomial data collected in a pilot study of a clinical trial.

```
data ptrials;
    input y n;
    datalines;
5 163
;
```

In the INPUT statement, Y represents the number of successes out of n independent Bernoulli trials. Similarly, for the current study, suppose you have the following hypothetical clinical trials data set BTRIALS.

```
data btrials;
    input y n @@;
    datalines;
2   86
2   69
1   71
1 113
1 103
;
```

In order to fit a model with a power prior, first combine both the current and pilot study data sets. The indicator GROUP variable takes on the values 'CURRENT' and 'PILOT' for the current and pilot study data set, respectively. The following SAS statements demonstrate how to create a new combined data set ALLDATA.

```
data alldata;
   set btrials(in=i) ptrials;
   if i then group='current';
   else group = 'pilot';
run;
```

Suppose you want to fit a Bayesian binomial model with a power prior for a hypothetical clinical trial setting with density

$$Y_i \quad \sim \quad \text{binomial}(n_i, p) \tag{3}$$

where p is the success probability for $i = 1, ..., 5$ observations.

The likelihood function for each of the observations is

$$p(Y_i | n_i, p) \quad = \quad \text{binomial}(n_i, p) \tag{4}$$

where $p(\cdot | \cdot)$ denotes a conditional probability mass function. The binomial density is evaluated at the specified value of Y_i and corresponding success probability p. Then, a power prior is used to assign each observation in the combined data set a binomial or weighted binomial likelihood function depending whether the observation was from the current or pilot study, respectively.

Suppose the following power prior is placed on the success probability parameters, where $L(p|Y_{0_i}, n_{0_i})$ represents the binomial likelihood:

$$
\begin{aligned}
\pi(p|D_0, a_0) &= L(p|Y_{0_i}, n_{0_i})^{a_0} \pi(p) \\
L(p|Y_{0_i}, n_{0_i}) &= \text{binomial}(n_{0_i}, p) \\
\pi(p) &= \text{uniform}(0, 1)
\end{aligned}
\tag{5}
$$

The variables Y_{0_i} and n_{0_i} represent the historical successes and sample size, respectively, and $\pi(\cdot)$ indicates a prior distribution. Suppose for this example that the discounting factor is set to $a_0 = 0.5$. The diffuse uniform$(0, 1)$ prior expresses your lack of knowledge about the success probability.

According to Bayes' theorem, the likelihood function and prior distribution determine the posterior distribution of p as given in Equation 2. PROC MCMC obtains samples from the desired posterior distribution, which is determined by the specified prior and likelihood. It does not require the exact form of the posterior distribution.

The following SAS statements use the likelihood and power prior distribution to fit the Bayesian binomial model with the historical data. The PROC MCMC statement invokes the procedure and specifies the input data set. The SEED= option specifies a seed for the random number generator, which guarantees the reproducibility of the random stream. The NMC= option specifies the number of posterior simulation iterations.

```
ods graphics on;
proc mcmc data=alldata seed=1181 nmc=10000;
   parms p 0.2;
   begincnst;
      a0 = 0.5;
   endcnst;
```

```
      prior p ~ uniform(0,1);
      llike = logpdf('binomial',y,p,n);
      if (group='pilot') then llike = a0*llike;
      model general(llike);
   run;
   ods graphics off;
```

The PARMS statement designates p as a parameter in the model and assigns an initial value of 0.2 to it. The programming statements enclosed by the BEGINCNST and ENDCNST statements are executed once per procedure call and designate the discount factor for the power prior.

The PRIOR statement specifies a prior p as given in Equation 5. The LLIKE= assignment statement assigns the log-likelihood function for X_i as given in Equation 3; the LOGPDF function calculates the logarithm of the binomial function. The IF/THEN statement uses the GROUP indicator variable to weight the binomial likelihood when the observation is from the pilot study.

The MODEL statement uses the GENERAL function, which indicates that you are using a SAS statement to construct a nonstandard density or distribution. The argument is an expression that takes the value of the logarithm function of the prior or likelihood distribution. In this case, the non-standard density is the binomial or weighted binomial likelihood, and the argument is the logarithm of the density.

Figure 1 displays convergence diagnostic plots for p to assess whether the Markov chains have converged. Inferences should not be made if the Markov chain has not converged.

Figure 1 Bayesian Binomial Model Diagnostic Plots for p
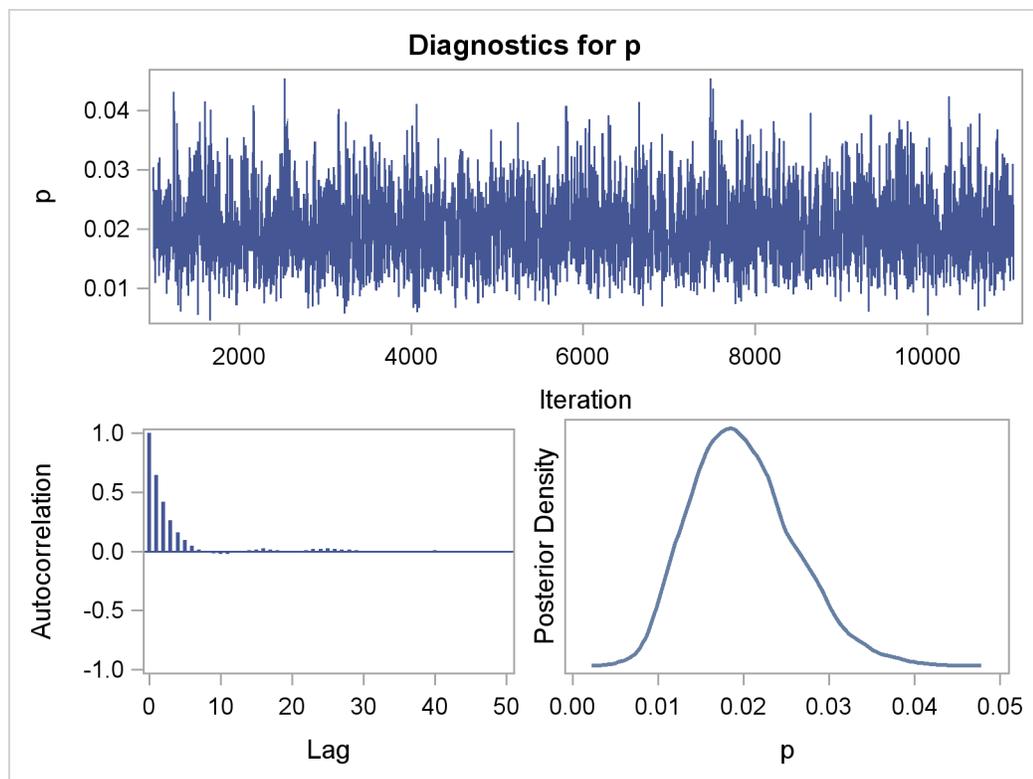
The trace plot shows that the mean of the Markov chain is constant over the graph and is stabilized. The chain is able to traverse the support of the target distribution, and the mixing is good. The trace plot shows that the Markov chain appears to have reached a stationary distribution.

The autocorrelation plot indicates low autocorrelation and efficient sampling. The kernel density plot shows a smooth, unimodal posterior marginal distribution for the parameter.

PROC MCMC produces formal diagnostic tests by default, but they are omitted here because informal checks on the chains, autocorrelation, and posterior density plots show desired stabilization and convergence.

Figure 2 reports summary and interval statistics for the parameter's posterior distribution.

Figure 2 Posterior Model Summary of Binomial Model

```
                         The MCMC Procedure

                        Posterior Summaries

                                 Standard            Percentiles
     Parameter         N     Mean    Deviation     25%       50%        75%

     p             10000   0.0200     0.00602   0.0156    0.0194     0.0237

                        Posterior Intervals

       Parameter    Alpha    Equal-Tail Interval        HPD Interval

       p            0.050      0.0100      0.0333    0.00900      0.0316
```

Using the power prior, the posterior mean of the success probability is 2.00% with a 95% credible interval of (1.00%, 3.33%), providing evidence of a treatment effect.

Modeling the clinical trial data set in a Bayesian binomial model with a power prior is a good way to make use of the pilot study. It enables you to use a different class of informative priors and incorporates relevant historical data and information.

References

Ibrahim, J. G. and Chen, M.-H. (2000), "Power Prior Distributions for Regression Models," *Statistical Science*, 15, 46–60.