August 27, 2016

Addition to PGLR Chap 6

Joseph M Hilbe, *Arizona State University*

Practical Guide to Logistic Regression
Ch 6: Bayesian Models for Count Data

Joseph M. Hilbe
11Sept, 2015: hilbe@asu.edu
Revised 2 Oct, 2015; revised 17 Nov, 2015

Bayesian modeling capabilities became available for researchers using Stata statistical software in April, 2015 with the release of Stata 14. Stata users now have a comprehensive platform to develop a truly wide range of Bayesian models. Moreover, the developers of **bayesmh**, Stata's core Bayesian modeling command, was designed it to provide users with ease of user together with maximum capabilities

6.2.4. Bayesian Logistic Regression Using Stata

Bayesian capabilities were added to Stata statistical software with version 14, which was released in April 2015. The final proofs pages of this book were submitted at the same time, but the production staff was able to include example code and partial output of a Bayesian logistic model using the new release of Stata. The code is on pages 148-149. This added section to Chapter 6 will expand a bit on Stata's Bayesian logistic regression capabilities.

The Stata *bayesmh* command provides a wide range of Bayesian modeling capabilities. The command allows Bayesian estimation using both built-in and user-defined non-informative and informative log-likelihood functions, as well as the same for built-in and user defined priors. The *logit* and *logistic* functions are used as a built-in or pre-defined canonical Bernoulli log-likelihood function, and the *binlogit* function is a pre-defined canonical binomial log-likelihood function. Recall that the canonical forms of the Bernoulli and binomial log-likelihoods are respectively employed for logistic and grouped logistic regression.

Stata also provides the user with a number of built-in univariate continuous prior distributions: normal, lognormal, uniform, gamma, inverse gamma, exponential, ebeta, chi2, jeffreys, and eight multivariate priors. Discrete priors offered include the Bernoulli, index and Poisson. Finally the flat prior is offered as a straightforward non-informative prior, and the logdensity prior is a generic prior which can be defined by the user. Note that the Cauchy and half-Cauchy priors, which are commonly used as priors for continuous predictors, do not have a built-in prior. Hence they must be defined by the user. An example will be provided in this section.

In the examples used for R's *MCMClogit* function and for the JAGS, logistic models, we used the *R84* (or *rwm1984*) data with outwork as the binary response variable and predictors of *female*, *kids, cdoc,* and *cage*. *Female* and *kids* are binary, and both *cdoc* and *cage* are centered versions of *docvis* and *age*. We shall use the same model here.

6.2.4.1: LOGISTIC REGRESSION NONINFORMATIVE PRIOR – USE BUILT-IN LOGIT

We first model the data using a standard logistic model.

```
. use rwm1984
. center docvis, pre(c)
. center age, pre(c)
. logit outwork female kids cdoc cage, nolog

Logistic regression                             Number of obs   =       3,874
                                                LR chi2(4)      =     1172.93
                                                Prob > chi2     =      0.0000
Log likelihood = -1959.1076                     Pseudo R2       =      0.2304
------------------------------------------------------------------------------
     outwork |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   2.256804   .0827624    27.27   0.000     2.094593    2.419016
        kids |   .3579763   .0899639     3.98   0.000     .1816503    .5343023
     cdocvis |   .0244323   .0062628     3.90   0.000     .0121574    .0367071
        cage |   .0543791   .0041591    13.07   0.000     .0462274    .0625308
       _cons |  -2.010276   .0810901   -24.79   0.000      -2.16921   -1.851342
------------------------------------------------------------------------------
```

The *bayesmh* command calls on an internal MCMC sampling procedure to produce posterior distributions for the four predictors.

```
. bayesmh outwork female kids cdoc cage,                      ///
              likelihood(logit) prior({outwork:}, flat) ///
              initial({outwork:} 0)

Bayesian logistic regression                    MCMC iterations  =      12,500
Random-walk Metropolis-Hastings sampling        Burn-in          =       2,500
                                                MCMC sample size =      10,000
                                                Number of obs    =       3,874
                                                Acceptance rate  =       .2062
                                                Efficiency:  min =      .02705
                                                             avg =      .04604
Log marginal likelihood = -1973.4741                         max =      .06151
------------------------------------------------------------------------------
             |                                                   Equal-tailed
     outwork |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
      female |  2.262614   .0825773   .004108   2.261485   2.104129   2.427077
        kids |  .3577974   .0875453    .00353   .3561315   .1882637   .5316499
     cdocvis |  .0243963   .0061635   .000375   .0242458   .0128949   .0372316
        cage |  .0544091   .0041393   .000173   .0543071   .0464286   .0625776
       _cons | -2.015416   .0802389   .003825   -2.01226  -2.179253  -1.857551
------------------------------------------------------------------------------
```

Notice the near identity of the standard frequency-based logistic regression and the Bayesian model with flat noninformative priors for all four predictors. The significance of the predictors are based on the credible intervals. None of the predictors have 0 within their credible interval. Therefore all are statistically significant.

The DIC statistic below is calculated as 3928.055. Since the DIC statistic, as AIC and BIC, is a comparative statistic, its value is not meaningful unless another model is being compared as to fit. Note also that the log-likelihood value is displayed as -1973. Compare this value to the standard model log-likelihood of -1959.

```
. bayesstats ic
```

```
Bayesian information criteria

-------------------------------------------------
             |        DIC     log(ML)     log(BF)
-------------+-----------------------------------
      active |   3928.055   -1973.474          .
-------------------------------------------------
Note: Marginal likelihood (ML) is computed
      using Laplace-Metropolis approximation.
```
Standard Bayesian graphs are provided as displayed below. A single command will display graphs of all four predictors and the intercept. I have chosen to display only the two continuous predictors.

```
. bayesgraph diagnostics {outwork: cdocvis}
```
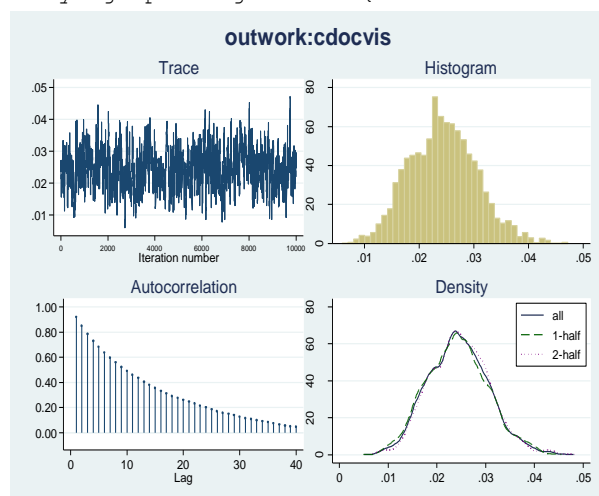

This is a diagnostic that a researcher would like to have for all studies. The trace in the upper left appears to display a fairly random iteration log. The histogram and density curve are fairly normal, and the mean values or estimates are close to the unimodal apex of the distribution. The graph of centered age is not as good as for doctor visits. If I were to run this model for a real research project, I would add more iterations to the sampling, and thin the selection of values by perhaps 3 or 5. Thinning has the sampling algorithm select every 5th observation to check, for example, not each observation. This helps prevent autocorrelation in the resultant posterior distribution. Note that the default number of sampling iterations we have been using for our example is 10,000. The default algorithm also employs a burn-in of 2500 iterations. Since early sampling is likely to be far off the true posterior mean value, a number of initial iterations are run, but do not figure in the calculation of the posterior mean. These initial iterations are referred to as *burnin()*. The default is 2500.

```
. bayesgraph diagnostics {outwork: cage}
```
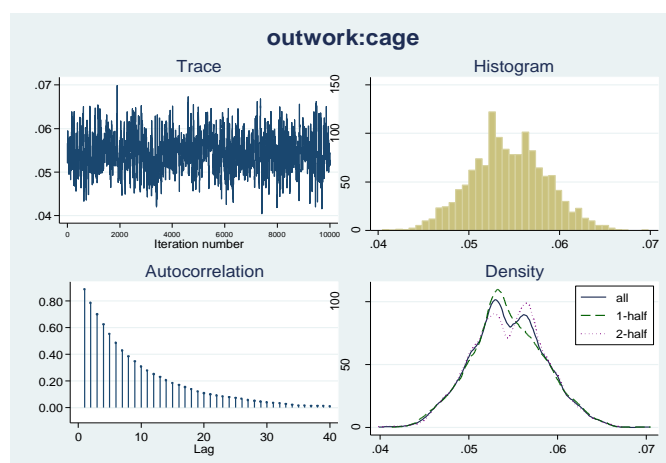
6.2.4.2: LOGISTIC REGRESSION NONINFORMATIVE PRIOR – USER DEFINED LOGIT

We may define our logistic log-likelihood function. For many – perhaps most -- models this is mandatory. I will provide an example of how to do this for the simple canonical Bernoulli (logistic) distribution, calling it *logitll*. The *llevaluator* function then calls this program.

I shall separate the log-likelihood function for when the response is 1 and when it is 0. Stata's programming code symbolizes response values as *$MH_y*.

LOGITLL LLevaluator Function

```
================================================
* Hilbe 2015
program define logitll
 version 14
 args lnf xb
 tempvar lnfj
 quietly generate `lnfj' = ln(invlogit( `xb')) if $MH_y == 1
 quietly replace `lnfj' = ln(invlogit(-`xb')) if $MH_y == 0
 quietly summarize `lnfj', meanonly
 if r(N) < $MH_n {
 scalar `lnf' = .
 exit
}
scalar `lnf' = r(sum)
end
=================================================
```

Now we model the data using the *logitll* program we just wrote. Note that we use a *llevaluator* function in place of a likelihood function.

```
. bayesmh outwork female kids cdocvis cage, llevaluator(logitll) prior({outwork:},
flat)


Bayesian regression                          MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling     Burn-in          =      2,500
                                             MCMC sample size =     10,000
                                             Number of obs    =      3,874
                                             Acceptance rate  =      .3377
                                             Efficiency:  min =      .0107
                                                          avg =     .03476
Log marginal likelihood = -1973.4765                      max =     .07626
```

```
-----------------------------------------------------------------------------
              |                                               Equal-tailed
      outwork |      Mean    Std. Dev.     MCSE     Median   [95% Cred. Interval]
--------------+--------------------------------------------------------------
       female |  2.259041    .0826589    .007992   2.257999   2.103572   2.422472
         kids |  .3561429    .0899495    .003257   .3553611   .1813112   .5314275
      cdocvis |  .0245713    .0063388    .000293   .0245858   .0121213   .0372038
         cage |  .0543688    .0041347    .000352   .0541872   .0466082   .0632439
        _cons | -2.011333    .0812332    .005014   -2.01019  -2.171695  -1.853979
-----------------------------------------------------------------------------
```

The posterior means – or Bayesian parameter estimates – are the same as we calculated for the standard logistic regression model and for the Bayesian logistic model with the built-in logit function. Below I give the DIC and log-likelihood value for this model. It is statistically identical to the previous calculation.

```
. bayesstats ic

Bayesian information criteria


-----------------------------------------------
             |     DIC     log(ML)     log(BF)
-------------+---------------------------------
      active |  3928.014   -1973.476          .
-----------------------------------------------
Note: Marginal likelihood (ML) is computed
      using Laplace-Metropolis approximation.
```

6.2.4.3: LOGISTIC REGRESSION INFORMATIVE PRIORS – USER DEFINED LOGIT

Now we shall display a model of the same data, but with half-Cauchy priors on the coefficients (parameters) of both *cdocvis* and *cage*. The half Cauchy is a weak informative distribution. We apply it to the standard deviation/variance of the predictors. Rather than give flat priors to the parameters for *female* and *kids*, we'll give what is perhaps better referred to as diffuse priors, which is close to non-informative. We'll give them a normally distributed prior, assuming that the posterior distribution of each is somewhat normally distributed --- which they are --- but gives them extremely wide variances. In fact, the variances are so wide that little information is given. If we were serious about this research, I would consider a beta prior for both the parameters for *female* and *kids*. But for this example a diffuse prior will do for the binary predictors. I will discuss more about using a beta prior for binary predictors following the example output.

In order to determine the half-Cauchy prior, we need to know the standard deviation of each of the continuous predictors. We use the summary command to obtain these values

```
. sum cdoc cage         # SDs  needed for bayesmh logdensity function

    Variable |       Obs         Mean    Std. Dev.         Min         Max
-------------+--------------------------------------------------------------
        cdoc |     3,874     7.82e-08    6.275955   -3.162881    117.8371
        cage |     3,874     1.22e-06     11.2401   -18.99587    20.00413
```

The half-Cauchy is defined here as (with the predictors symbolized as X)

$$\log(sd(X)) - \log(sd(X)^2 + X^2) - \log(\pi)$$

Note that the half-Cauchy prior specification from the logistic model may be used in other models such as Poisson. The general specification is

```
prior({outwork:cage}, logdensity(ln(11.24)-ln(11.24^2+({outwork:cage})^2)-ln(_pi)))
```

regardless of the model being used. Priors are relevant to the likelihood of the predictor.
   Alternatively, you may first rescale the `cage` variable by dividing it by its standard deviation as
```
        . summarize cage. replace cage = cage/`r(sd)'
```

and then by applying the standard Cauchy distribution (with scale 1) as the prior

```
    prior({outwork:cage}, logdensity(-ln(_pi*(1+({outwork:cage})^2))))
```

The half-Cauchy as defined above will be used in the log-likelihood function below.

```
============================================================
bayesmh outwork female kids cdocvis cage, llevaluator(logitll)        ///
    prior({outwork:female kids _cons}, normal(0, 100000))             ///
    prior({outwork:cdocvis},                                          ///
    logdensity(ln(6.276)-ln(6.276^2+({outwork:cdocvis})^2)-ln(_pi)))  ///
    prior({outwork:cage},                                             ///
    logdensity(ln(11.24)-ln(11.24^2+({outwork:cage})^2)-ln(_pi)))     ///
    block({outwork:female kids _cons})
============================================================
```

```
Bayesian regression                         MCMC iterations   =    12,500
Random-walk Metropolis-Hastings sampling    Burn-in           =     2,500
                                            MCMC sample size  =    10,000
                                            Number of obs     =     3,874
                                            Acceptance rate   =     .2586
                                            Efficiency:  min  =    .05607
                                                         avg  =    .07764
Log marginal likelihood =  -27363.58                     max  =     .1156
------------------------------------------------------------------------------
             |                                              Equal-tailed
     outwork |      Mean    Std. Dev.     MCSE      Median   [95% Cred. Interval]
-------------+----------------------------------------------------------------
      female |  2.241821    .0832145   .002934    2.242154   2.085017   2.412527
        kids |  .2955727    .0897766   .003791    .2956706   .1274696   .4799008
     cdocvis |   .019811    .0054574   .000161    .0198157   .0091171   .0302249
        cage |  .0486332    .0037328   .000145    .0484574   .0414594   .0560532
       _cons | -1.966496    .0796124   .003023   -1.964917  -2.122419  -1.816811
------------------------------------------------------------------------------
```

```
. bayesstats ic

Bayesian information criteria

-------------------------------------------------
             |       DIC     log(ML)     log(BF)
-------------+-----------------------------------
      active |  3930.211   -27363.58           .
-------------------------------------------------
Note: Marginal likelihood (ML) is computed
      using Laplace-Metropolis approximation.
```

Regarding the use of a beta prior for binary predictors: this is closely related to prior elicitation. Suppose that you are interested in eliciting a beta-prior for the probability {p} of being a *female*; i.e., you want to specify a Beta(a,b) prior for {p} for some parameters -a- and -b-. One way to determine -a- and -b- is to match the first two moments on the beta-distribution. You will need *a priori* information about the expected {p} in the general population. Suppose *mu*=0.51, and its variance, *sig2*=0.051 (SD = 0.2258). *sig2* estimates the strength of your belief in *mu*: the smaller value for *sig2,* the stronger the beliefs. Then you have the equations

```
    a = (mu*(1-mu)/sig2 - 1)*mu     = 1.989

. di (0.51*(1 - 0.51)/0.051 - 1)* 0.51
1.989

    b = (mu*(1-mu)/sig2 - 1)*(1-mu) = 1.911

. di (0.51*(1 - 0.51)/0.051 - 1)* (1 - 0.51)
1.911
```

You may then specify -*prior({p}, Beta(1.989, 1.911))*- to express your *a priori* information about {p}. This is an informative prior, but not a very strong one. For example, Beta(19.89,19.11) would have the same expectation of 0.51 for {p} but would be much stronger.
   Finally, we'll obtain diagnostic graphs of the parameters for the two continuous predictors. We'll be particularly interested if the graph of the parameter of *cage* is better distributed than before. If so, the half-Cauchy prior is likely a cause.

. bayesgraph diagnostics {outwork: cdoc}
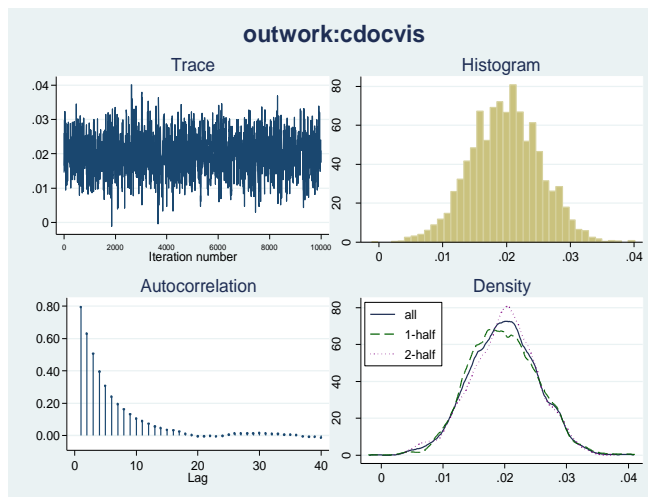
. bayesgraph diagnostics {outwork; cage)
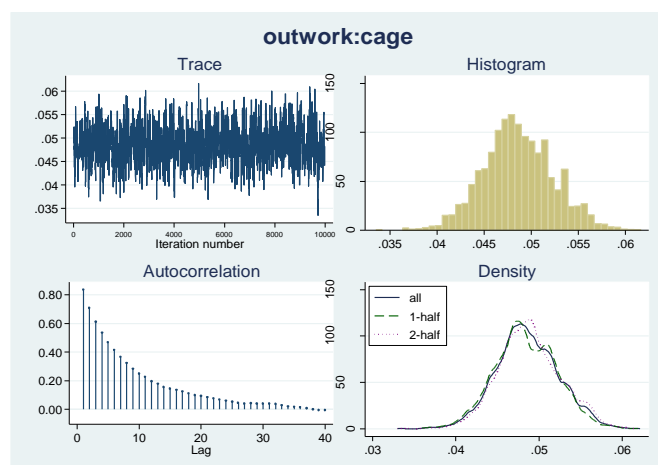
This graph is much better distributed than before. The earlier graph was nearly bimodal. Note that the credible interval of *cage* is more narrow with this model than for the models with non-informative priors. The prior seems to have helped effect a better fit for centered age. In fact, the interval is more narrow for *cdocvis* with the prior than without it.

6.2.4.4 BAYESIAN GROUPED LOGISTIC REGRESSION

Grouped logistic regression models were discussed in the previous chapter. Here we model grouped data using Bayesian methodology. Fortunately Stata has provided a built-in function for the binomial log-likelihood called *binlogit*. The function comes with an option, or argument, identifying the variable in the data that defines groups.

Recall that each observation in a grouped model has a response variable consisting of two components: a denominator specifying the number of observations with predictors having the same values, and a numerator depicting the number of cases in each observation for which the response variable has a value of 1.

We'll use the 1912 Titanic shipping disaster data in grouped format for an example. The variable *cases* tells us the number of observations having an identical pattern of covariate values. The variable *survive* tells us how many people in each covariate pattern survived. With binary variables *age* and *sex* and a three level categorical predictor, *class,* as predictors, the model may be set using the code below. A flat prior is given each predictor parameter estimate.

```
. use titanicgrp, clear
. bayesmh survive i.age i.sex i.class, likelihood(binlogit(cases)) ///
        prior({survive:}, flat)

.   .   .

Bayesian binomial regression                  MCMC iterations  =    12,500
Random-walk Metropolis-Hastings sampling      Burn-in          =     2,500
                                              MCMC sample size =    10,000
                                              Number of obs    =        12
                                              Acceptance rate  =     .2727
                                              Efficiency:  min =    .03507
                                                           avg =    .04823
Log marginal likelihood = -79.019375                       max =    .08408
```

```
-------------------------------------------------------------------------------
                 |                                              Equal-tailed
       survive   |      Mean    Std. Dev.      MCSE     Median   [95% Cred. Interval]
-----------------+-------------------------------------------------------------
           age   |
        adults   |  -1.076341    .2428045    .012413  -1.074503  -1.577094   -.5995606
                 |
           sex   |
           man   |  -2.386441    .1461116    .007802  -2.385702  -2.680732   -2.099317
                 |
         class   |
     2nd class   |  -1.025378    .1929263    .006653  -1.033393  -1.399935    -.657782
     3rd class   |  -1.787636    .1734327    .007926  -1.785742  -2.141809   -1.458485
                 |
         _cons   |   3.105963    .3047856     .01609   3.099388    2.52093    3.729154
-------------------------------------------------------------------------------
```

We shall compare the above Bayesian grouped logistic model with a standard grouped logistic regression. We expect that the mean values for the posterior distribution of each predictor are nearly the same as the parameter estimates in the standard model. For space purposes the header statistics will not be displayed.

```
. glm survive i.age i.sex i.class, fam(bin cases) nolog nohead

-------------------------------------------------------------------------------
                 |                 OIM
       survive   |      Coef.    Std. Err.       z    P>|z|      [95% Conf. Interval]
-----------------+-------------------------------------------------------------
           age   |
        adults   |  -1.055608    .2426561    -4.35   0.000    -1.531205   -.5800106
                 |
           sex   |
           man   |  -2.369465    .1452512   -16.31   0.000    -2.654152   -2.084778
                 |
         class   |
     2nd class   |  -1.010558    .1949348    -5.18   0.000    -1.392623   -.6284927
     3rd class   |  -1.766372    .1707174   -10.35   0.000    -2.100971   -1.431772
                 |
         _cons   |   3.061882    .2980054    10.27   0.000     2.477802    3.645962
-------------------------------------------------------------------------------
```

The values are close, but can be made closer with more iterations, and thinning. Adding the following line to the code increases the default number of samples from 10,000 to 40,000, a burn-in of 10,000 in place of the default 2,500 and thinning of 3. The default value for thinning is 1, which causes every sample draw to be checked. Using these amendments should   bring the Bayesian (noninformative) and standard frequentist model values closer. More iterations and perhaps a longer burnin could bring the values even closer.

```
. bayesmh survive i.age i.sex i.class, likelihood(binlogit(cases)) ///
        prior({survive:}, flat)  ///
        mcmcsize(40000) burnin(10000) thinning(3)

Bayesian binomial regression                      MCMC iterations  =     129,998
Random-walk Metropolis-Hastings sampling          Burn-in          =      10,000
                                                  MCMC sample size =      40,000
                                                  Number of obs    =          12
                                                  Acceptance rate  =       .1547
                                                  Efficiency:  min =      .04913
                                                               avg =       .1105
Log marginal likelihood = -79.00091                            max =       .1614
```

```
------------------------------------------------------------------------
             |                                             Equal-tailed
     survive |      Mean    Std. Dev.      MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------
         age |
      adults | -1.062168    .2437522     .00335   -1.06257  -1.544836  -.5845023
             |
         sex |
         man | -2.380523    .1417244    .003197  -2.379758  -2.659573  -2.107879
             |
       class |
   2nd class | -1.016803    .1949742    .002427  -1.015071  -1.401847  -.6354437
   3rd class |  -1.77442    .1708537    .002562  -1.772023  -2.117457  -1.446969
             |
       _cons |  3.078187    .2990453    .004769   3.075592   2.498903   3.678929
------------------------------------------------------------------------
```

Care must be taken in assigning priors to the coefficients of model predictors. There is a lot of literature on the subject, but it extends beyond the scope of our discussion. Bayesian modeling is becoming very popular in many areas of research. In astrostatistics, for example, most all new statistical analysis of astrophysical data entails Bayesian modeling. Ecologists are also basing more and more of their research on Bayesian methods. Researchers in other areas are beginning to do the same. I have just touched on the basics of Bayesian modeling in this chapter.