Models In Epidemiology And Biostatistics
Gordon Hilton Fick

An Introduction To Bayesian Methods

Some Discussion :

Bayes Rule is included in most introductory material on Probability. We have seen Bayes Rule in the material in Session 9 on Sensitivity, Specificity and Prevalence leading to determinations of Predictive Value Positive and Predictive Value Negative.
Sackett [1985] contains a fascinating tour through the effective use of these matters.
We then have very clearly defined prior probabilities [the pre-test probabilities] and the application of Bayes Rule is rarely in dispute.

Until the 1950's, Bayesian methods were rarely seen in Statistics. The cases for and against Bayesian statistics have been stated and argued ever since. It remains unclear whether this controversy will ever be resolved.

In current times, for many Biostatisticians and Epidemiologists, the controversies are of no interest and there is a view close to pragmatism in vogue. In some settings, Bayesian analyses give numbers that are not materially different from Classical analyses. By Classical, it is meant, here, that the analyses do not come from the Bayesian paradigm. The distinction between Classical and Bayesian then dominated by interpretation of the results.

There are now a number of new analysis methods that come from a Bayesian approach and there is no easy way to see these new methods as being Classical.

For this session, I will begin with the Bayesian approach to Logistic Regression. The algorithms to carry out this approach are now included in recent releases of Stata and in the form of many R packages.

There are now many books entirely devoted to Bayesian methods. There are a number of books in Epidemiology or Biostatistics which include at least one chapter on Bayesian methods.

I have always appreciated the commentary in Fraser[1976]. I have included it with this session.

An Example :

Let us return to the Kalbfleisch data introduced in the very first session and consider a Bayesian approach. You may want to have another look at the original details in Session 1.

Here are the Classical results with the logit link.

```
. use kalbfleisch.dta

. gen s=surg-1

. gen st=s*tr
. cc suc tr,by(surg)

            surg |       OR       [95% Conf. Interval]   M-H Weight
```

```
-----------------+------------------------------------------------
               1 |   2.111111      .8436801   6.804342     4.090909 (exact)
               2 |         19      7.753503   60.26036     2.272727 (exact)
-----------------+------------------------------------------------
           Crude |   .2538701      .2077658   .3099697              (exact)
    M-H combined |   8.142857      4.342777   15.26814
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    11.57  Pr>chi2 = 0.0007

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     67.85
                                                Pr>chi2 =    0.0000

. logit suc tr s st

Logistic regression                             Number of obs     =      2,200
                                                LR chi2(3)        =     636.30
                                                Prob > chi2       =     0.0000
Log likelihood = -1057.9332                     Pseudo R2         =     0.2312

------------------------------------------------------------------------------
         suc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |   .7472144   .4707838     1.59   0.112     -.175505    1.669934
           s |   2.944439   .4631699     6.36   0.000     2.036643    3.852235
          st |   2.197225   .6604269     3.33   0.001     .9028116    3.491638
       _cons |  -2.944439   .4588315    -6.42   0.000    -3.843732   -2.045146
------------------------------------------------------------------------------

. disp exp(0.7472144)
2.1111111

. disp exp(2.197225)
9.0000038
```

Logistic Regression [with logit] reproduces the Stratified Analysis [with cc]. Logistic Regression works with the log-likelihood function by determining the maximum of this function and then standard errors are based on the curvature of the log-likelihood at the maximum.

Bayesian methods use an 'extended' version of Bayes Rule :
"The posterior distribution is proportional to the product of the prior distribution and the likelihood function" :

$$p(\boldsymbol{\beta}:\boldsymbol{y}) \propto p(\boldsymbol{\beta}) * L(\boldsymbol{y}|\boldsymbol{\beta})$$

One then seeks to determine a prior that incorporates the "knowledge" about the parameter $\boldsymbol{\beta}$ in the form of a probability distribution.
Then, one usually determines the marginal posterior distribution for each component of $\boldsymbol{\beta}$:

$$p_i(\beta_i:\boldsymbol{y})$$

Characteristics of these marginal posteriors are then reported. One typically sees the reporting of the mean, the median and the standard deviation of each marginal posterior.
Also, one sees what have come to be called the "credible intervals". A 95% credible interval is usually determined from 2.5% point and the 97.5% point of each marginal posterior.
Credible intervals offer a direct probability statement without the "repeated sampling" phrase attached with confidence intervals.

Algorithms are now available to determine the posterior, the marginal posteriors and the characteristics of the marginal posteriors. The algorithms are "computationally intensive". Most users of such algorithms review many ways to check that the results are "numerically correct".
The development of these algorithms has been underway for many years and advances continue to this

day. There are now many books devoted almost entirely to these algorithms.

A Noninformative Prior :

It is of interest to explore when Bayesian numerical results will be close to the Frequentist numerical results. It can be noticed that if a prior is constant [or "flat"] then the posterior is proportional to the likelihood. Notice that, in the case of a flat prior, the mode of the posterior will then be the Maximum Likelihood estimate. Here, we are referring to the joint posterior and not the individual marginal posteriors.

So let us try the Bayesian model :

$$\log\left(\frac{p}{1-p}\right)=\beta_0+\beta_1 T+\beta_2 S+\beta_3 ST \quad p(\boldsymbol{\beta})=c$$

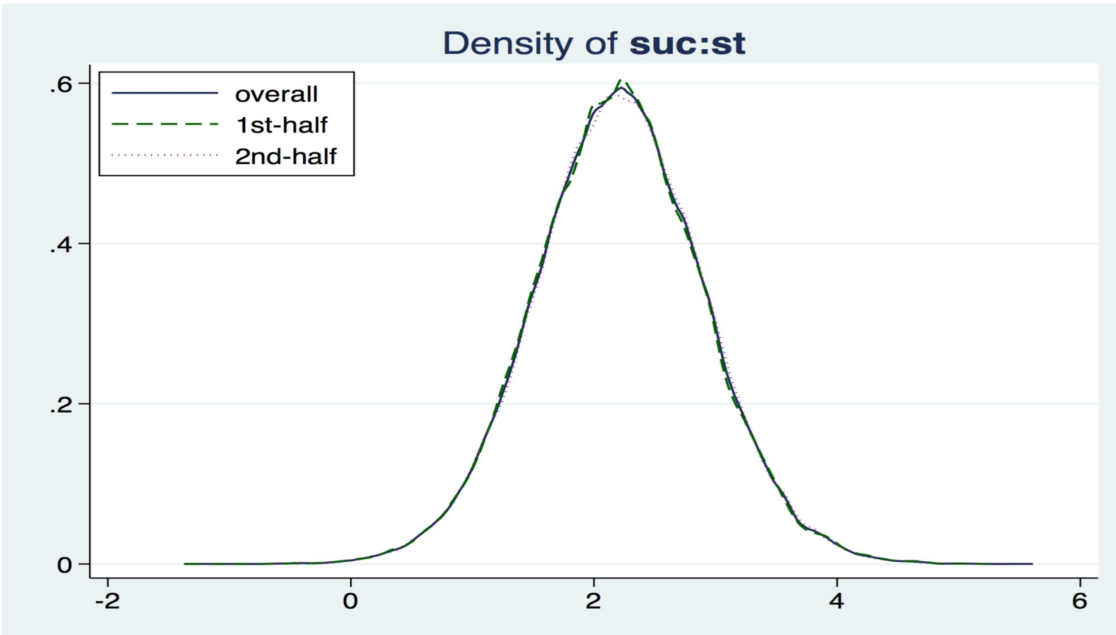
```
. bayesmh suc tr s st, likelihood(logit) prior({suc:},flat)

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  suc ~ logit(xb_suc)

Prior:
  {suc:tr s st _cons} ~ 1 (flat)                                           (1)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_suc.

Bayesian logistic regression                     MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      2,200
                                                 Acceptance rate  =      .2061
                                                 Efficiency:  min =     .05507
                                                              avg =     .06382
Log marginal likelihood = -1060.7756                          max =     .07996

------------------------------------------------------------------------------
             |                                                Equal-tailed
         suc |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
          tr |  .8469676   .4960566   .019813   .8391533  -.0485705    1.88898
           s |  3.054957   .4878875   .020337   3.032616   2.158087     4.0481
          st |  2.205466   .7052731   .024941    2.19165   .8109429   3.676366
       _cons | -3.053498   .4808389   .020489  -3.040387  -4.078676  -2.182822
------------------------------------------------------------------------------
```

From this model, our interpretation would begin with number(s) listed for β_3 which is `2.205466` for the mean of the marginal posterior and is `2.19165` for the median of the marginal posterior.
Note that the MLE was `2.197725` which is the mode from the [joint] posterior here.
All of the numbers listed in the table above are computed from the marginal posteriors. The exception is the heading MCSE which provides a measure of the numerical accuracy of the algorithm.

Notice that, with this Bayesian approach, there are no tests of significance and estimation, per se, is not relevant. The result of a Bayesian analysis is the posterior distribution and the marginal posterior distributions. We can display a graph of each of the marginal posteriors. In principle, we could display contour plots of for posteriors marginal with respect to two of the parameters.

Here, for example, is the marginal posterior for β_3 :
Using :

```
. bayesgraph kdens {suc:st}
```



A Bayesian analyst would say that all inferences for β_3 would be obtained from this display and probability calculations are then obtained from it.

An Informative Prior :

We now consider another analysis using an informative prior. The Bayesian model can specify a wide range of different priors.
As an example only, let us suppose that we are apriori "fairly sure" that surg does not modify. We could use a Normal(0,0.25) prior for st rather than removing st from the model.

```
. bayesmh suc tr s st, likelihood(logit) prior({suc:st},normal(0,0.25)) prior({suc:tr s
_cons},flat)

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  suc ~ logit(xb_suc)

Priors:
          {suc:st} ~ normal(0,0.25)                                          (1)
  {suc:tr s _cons} ~ 1 (flat)                                                (1)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_suc.

Bayesian logistic regression                     MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
```

```
                                                  Number of obs    =      2,200
                                                  Acceptance rate  =      .1881
                                                  Efficiency:  min =     .06784
                                                               avg =     .07285
Log marginal likelihood = -1064.8049                           max =     .07506

------------------------------------------------------------------------------
             |                                                Equal-tailed
         suc |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
          tr |  1.784609   .4801246   .017616   1.772891    .873322   2.773021
           s |  3.959096   .4797938   .017513   3.950082   3.019005   4.932413
          st |  .7193152   .4349067   .016698   .7305172  -.1792441   1.542511
       _cons | -3.950332   .4795938   .017604  -3.926571  -4.934509  -3.010152
------------------------------------------------------------------------------
```

Improper Priors :

You will have noticed that the constant [or flat] prior is not a probability distribution. Such priors are often called improper priors. There is considerable debate as to whether improper priors should ever be used. Rather than using so-called non-informative priors, some authors advocate for the use of "weakly informative" priors. For example, if one considers a Normal distribution with a huge variance then such a prior is integrable and is "proper".

There is wide range of models where the use of a flat prior is definitely acceptable. For example, Fraser [1976] Chapter 11 shows that Linear Regression can be presented with a conditioning argument that is most definitely correct and within the Classical world. The numerical results with Fraser's approach are identical to those that would be obtained from the Bayes [with a flat prior] approach. The interpretations are not the same : Fraser's approach gives Classical notions like tests of significance and confidence intervals.

Full disclosure : GHF has contributed to this literature using Fraser's approach.

Independent Priors :

Implicit in our above examples is the assumption of independence for the components of the prior. There are many circumstances where an analyst will know that components of a prior cannot be independent. One example would be the coefficients for slope and intercept with a measured [uncentred] variable like age. One is then obliged to use priors that are multivariate distributions with builtin correlations.

The Algorithms :

In recent years, there have been major advances in the algorithms for the computing in Bayesian methods. In particular, the Metropolis-Hastings [M-H] algorithms have shown substantial gains. Further, one gets Markov Chain Monte Carlo [MCMC] results and measures of the accuracy of the results [MCSE, for example]. Nevertheless, the more complex the model, the closer one is to the frontier of computing viability. There are now now a number of measures and graphics available to assess convergence issues and viability issues. Interpretation challenges abound.

Empirical Bayes :

In Session 22 and later, we have subject specific components $\boldsymbol{u}$ in the conditional models. In this framework, these unknowns really do have probability distributions. In the simplest form, the probability distribution might have mean zero and $VAR\ \boldsymbol{u} = \sigma_{\boldsymbol{u}}^2 I$. In Session 22, there is mention of the determining predictions for the $\boldsymbol{u}$. It turns out that the most widely used method for constructing these predictions is a hybrid of Classical and Bayesian called Empirical Bayes.

Our approach to the study of the conditional models is Classical for the parameters $\boldsymbol{\beta}$ and σ_u^2 and is based on the likelihood function. One would determine the MLE for $\boldsymbol{\beta}$ and σ_u^2 and offer standard errors and confidence intervals as usual.

Now since the $\boldsymbol{u}$ have a distribution we can conceptualize a posterior for $\boldsymbol{u}$ as :

$$p(\boldsymbol{u} : \boldsymbol{y} \mid \boldsymbol{\beta}\ \ \sigma_{\boldsymbol{u}}^2) \ \propto\ p(\boldsymbol{u} \mid \sigma_{\boldsymbol{u}}^2) * L(\boldsymbol{y} \mid \boldsymbol{\beta}\ \ \sigma_{\boldsymbol{u}}^2)$$

where the prior distribution is that distribution with mean zero and $VAR\ \boldsymbol{u} = \sigma_{\boldsymbol{u}}^2 I$

So the posterior for $\boldsymbol{u}$ is a function of $\boldsymbol{\beta}$ and σ_u^2 . Now, we replace the unknowns $\boldsymbol{\beta}$ and σ_u^2 with their MLEs in this posterior for $\boldsymbol{u}$. This is the Empirical Bayes step. This computed posterior can then be studied. One can determine characteristics of the marginal posteriors as above. One usually sees the means of these marginal posteriors reported. These means are the predictions of the $\boldsymbol{u}$.

Stata computes these predictions using a postestimation predict command :

```
. predict u, reffects
```

as in Session 22.

Using a Bayesian model when a Classical model fails:

There are articles where the author presents a Bayesian analysis with the remark that the Classical approach failed in some way. Such an attitude to abandoning a Classical analysis could be most unsatisfactory. For example, suppose a Classical analysis results in a variable deletion because of "multicollinearity" while the Bayesian analysis leaves the variable in the results. One needs to be very sure that the model is formulated correctly and certain logic errors are not present in the model development. The Bayesian analysis may simply be hiding a more serious issue.

Bayesian models that do not have Classical counterparts or where Classical approaches are currently intractable :

There are now many Bayesian models that are often very complex but the analyses resulting from these models are available. This literature is advancing and is quite interesting. The joint posteriors and marginal posteriors do need substantial checking. The numerical issues in play can be concerning. Further, there is a literature offering conditions for when such computed probabilities are real or have issues regarding inconsistencies.

Conjugate Priors :
There is a large literature on priors that make the mathematics possible. One can then get so-called closed form expressions for the posteriors. These priors are usually called 'conjugate'. Conjugate, as an adjective, means 'joined in pairs'. Really, these priors provide some entertaining mathematics and,

perhaps, some guidance to the implications to the choice of supposed 'real' priors. Further, the circumstances where any type of conjugate prior is available tend to be for very simple settings only.

Further reading :
This session provides only a brief [and arguably] superficial introduction to Bayes. This literature is changing rapidly. Older papers and books may be essentially obsolete or providing only a historical take on this topic.

Lesaffre E Lawson AB 'Bayesian Biostatistics' [2012]
Berry DA Stangl DK 'Bayesian Biostatistics' [1996]