

Teaching Bayesian Statistics in Undergraduate Classes

Ananda Jayawardhana

Pittsburg State University

Pittsburg, KS 66762

Abstract: Bayesian statistics is just another chapter at the end of the semester in mathematical statistics sequences for which much time is not allocated. Students consider it as another section of statistics rather than a totally different way to look at the entire field of statistics. At lower level undergraduate classes, students do not even hear about Bayesian statistics. In this paper, the author will discuss his effort to introduce Bayesian concepts to several statistics classes at undergraduate level. Difficulties of teaching for the instructor and difficulties of understanding for the students will also be discussed.

Key Words:

1. Introduction

Bayesian statistics has come a long way over the last several decades. There are free software such as WinBUGS available to easily implement Bayesian methods. World scientific community has started recognizing Bayesian methodologies. With this development, it is advantageous for modern graduate to know Bayesian methods as well as frequentist methods. With the availability of new software it is not very difficult to add some Bayesian concepts to lower level statistics classes. This paper will discuss how to incorporate Bayesian methods in undergraduate classes. Few lectures of applied Bayesian methods can be easily added to existing applied statistics courses. Some theoretical background can be added to typical mathematical statistics courses with programming component. Instructors can encourage undergraduate students to do honors projects using Bayesian methods, especially with computing. One way to get the attention of the students is to include a few lectures in Bayesian methods with more emphasis in computing in undergraduate applied statistics classes. Theory and applied Bayesian methods can be easily added to mathematical statistics sequence. Some textbook authors such as Rice, J. A. (2009) have added Bayesian ideas in appropriate places of their books. Finally, it is easy to add Bayesian methods for upper level undergraduate/ graduate level applied statistics courses such as Applied Linear Regression. An instructor who teaches undergraduate statistics courses can make some desired outcomes such as ability to find credible intervals for mean and proportion, conduct Bayesian hypotheses testing for mean and proportion, and conduct Bayesian linear regression. For some undergraduate students who take graduate level courses, one can expect goals of higher level such as ability to conduct logistic regression using Bayesian methods, select reasonable priors, find theoretical posterior distributions, prior, and posterior predictive distributions, write WinBUGS and R codes and understand basics of MCMC methods.

In this paper coding in WinBUGS is used but not explained. Readers are expected to know WinBUGS. Learning WinBUGS for an instructor is an easy task.

There are a few articles written about teaching Bayesian statistics at lower level or service statistics classes. Albert, J. H. (1993) discusses the difficulties of deriving posterior distributions and propose simulation approach to posterior distributions. Those of who have tried simulation of posterior distributions prior to 1993 may remember the difficulties involved with such simulations. He gives an example of baseball home run prediction which will be revisited in this paper using modern Bayesian

software. Berry (1997), discusses the difficulties of teaching Bayesian statistics at elementary statistic level and provides a list of topics one can cover if such a course is planned. In his article, he refers to some Minitab macros written by James Albert in 1996 as some examples to introduce Bayesian statistical software to students. In my classes students found a lot of online resources but I recommended Lunn et. al. (2013) “The BUGS Book” as a reference.

2. Getting the Attention of Students

Using simulation methods to approximate integrals is one way to get the attention of undergraduate mathematics students. For other students, rather than plugging numbers in to an equation to find quantities such as mean and variance of distributions, simulations are an attractive alternative.

Example 1: Evaluation of the integral of x^3 from 0 to 1 yields 0.25. Using the WinBUGS code:

```
model{
Integral~<- pow(x,3)
x~dunif(0,1)
}
```

and calculating the mean of the random variable *Integral* will produce an approximate value. See Table 1.



| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|----------|--------|--------|----------|----------|--------|-----------|-------|--------|
| Integral | 0.2493 | 0.2819 | 0.002798 | 2.024E-5 | 0.126 | 0.9279 | 1 | 10000 |

Table 1.

Example 2: Evaluate the mean of the exponential distribution with parameter 2. Integration will yield a value of 0.5.

Using the WinBUGS code:

```
model{
x~dexp(2)
}
```

and calculating the mean of the random variable x will produce an approximate value. See Table 2.



| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|---|--------|--------|----------|----------|--------|-----------|-------|--------|
| x | 0.5002 | 0.4987 | 0.002273 | 0.01267 | 0.3476 | 1.835 | 1 | 50000 |

Table 2.

At this stage it is possible to explain that the expected value of a function $g(x)$ can be approximated by $E[g(x)] = \frac{1}{N} \sum_{i=1}^N g(x_i)$ where random values of x_i are generated from the desired distribution.

Example 3:

Probability integral transform. Let $F(x) = 1 - \exp(-x)$, then $F(x) \sim \text{Uni}(0, 1)$ and $x = -\log(1 - F(x)) = -\log(u)$.

Using the WinBUGS code:

```
model{
u~dunif(0,1)
x<- -log(u)
}
```

and plotting the posterior densities, one can show that the distribution of x is close to exponential. See Figure 1.

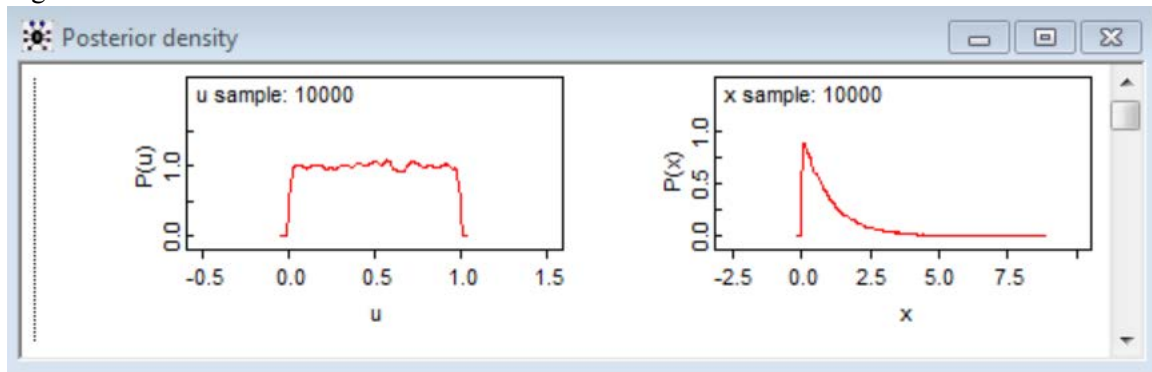


Figure 1.

Example 4: Probability distribution tables.

Let $X \sim \text{Bin}(n = 4, p = 0.4)$ and we want to approximate the cumulative distribution function table.

Using the WinBUGS code:

```
model{
x~dbin(0.4,5)
cdf0<-1-step(x-0.5)
cdf1<-1-step(x-1.5)
cdf2<-1-step(x-2.5)
cdf3<-1-step(x-3.5)
cdf4<-1-step(x-4.5)
cdf5<-1-step(x-5.5)
}
```

and calculating the mean of the random variables one can approximate the cdf values for $x = 1, 2, 3, 4, 5$. See Table 3. If the summary statistics of the node x was requested one can get the mean of the distribution too

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|------|--------|--------|----------|----------|--------|-----------|-------|--------|
| cdf0 | 0.0818 | 0.2741 | 0.00263 | 0.0 | 0.0 | 1.0 | 1 | 10000 |
| cdf1 | 0.3439 | 0.475 | 0.004254 | 0.0 | 0.0 | 1.0 | 1 | 10000 |
| cdf2 | 0.6766 | 0.4678 | 0.004147 | 0.0 | 1.0 | 1.0 | 1 | 10000 |
| cdf3 | 0.9096 | 0.2868 | 0.002835 | 0.0 | 1.0 | 1.0 | 1 | 10000 |
| cdf4 | 0.9892 | 0.1034 | 0.001161 | 1.0 | 1.0 | 1.0 | 1 | 10000 |
| cdf5 | 1.0 | 0.0 | 1.0E-12 | 1.0 | 1.0 | 1.0 | 1 | 10000 |

Table 3.

Example 5: Transforming normal distributions to standard normal distributions.

Using the WinBUGS code:

```

model{
x~dnorm(50,0.25)
z<-(x-50)/2
}
    
```

and graphing the random variables x and z one can approximate the distributions. Also summary statistics provide approximations to the means and standard deviations. See Figure 2 and Table 4.

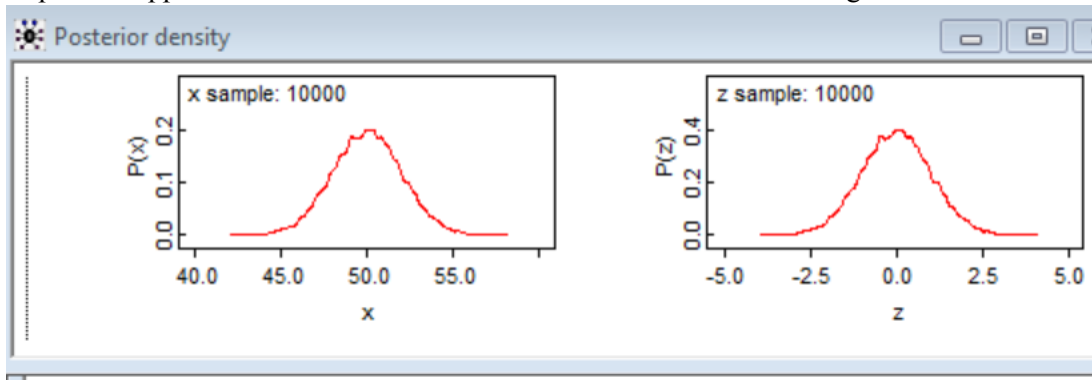


Figure 2.

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|---|-----------|-------|----------|----------|---------|-----------|-------|--------|
| x | 50.0 | 2.01 | 0.02181 | 46.09 | 50.01 | 53.98 | 1 | 10000 |
| z | -0.001535 | 1.005 | 0.0109 | -1.954 | 0.00317 | 1.989 | 1 | 10000 |

Table 4.

3. Introducing Bayesian Concepts

Example 6: Using the data 1192, 1200, 1207, 1185, 1198, 1194, 1210, 1197, 1212, 1209, 1189, 1202, 1194, 1196, 1179, 1191, 1214, 1197, 1213, 1213, 1211, 1193, 1204, 1187, 1196, 1194, 1220, 1193, 1194, 1194, 1177, 1181, 1217, 1213, 1204, 1119, 1221, 1198, 1210, 1183, find a 95% confidence interval for the mean. Assume the data have a normal distribution with standard deviation of 12. Kokoska, S (2015)

The frequentist confidence interval is $1199.47 \pm 1.96 \left(\frac{12}{\sqrt{40}} \right)$ which produces [1195.75, 1203.19].

Using the WinBUGS code:

```
model{
for (i in 1:40)
{
x[i]~dnorm(mu,tau) #Model
}
mu~dunif(1000,1500) #Prior distribution for mu
tau<-1/144
}
#Data
list( x=c(1192, 1200, 1207, 1185, 1198, 1194, 1210, 1197, 1212, 1209, 1189, 1202, 1194, 1196,
1179, 1191, 1214, 1197, 1213, 1213,
1211, 1193, 1204, 1187, 1196, 1194, 1220, 1193, 1194, 1194, 1177, 1181, 1217, 1213, 1204,
1119, 1221, 1198, 1210, 1183))
#Initial values
list(mu=1100)
and calculating the summary statistics, a 95% credible interval is [1194,1201]. Students should
be encourages to use different priors and how the credible interval changes. See Table 5.
```

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|----|--------|-------|----------|----------|--------|-----------|-------|--------|
| mu | 1198.0 | 1.907 | 0.02066 | 1194.0 | 1198.0 | 1201.0 | 1 | 10000 |

Table 5.

Example 7: Using the data 3.4, 3.5, 3, 3.4, 3.3, 3.5, 3.3, 3, 3.5, 3, 2.7, 3.1, 3.8, 3.1, 3.3, 3.1, 3.1, 2.9, 2.5, 3.1, 3, 3.6, 3.2, 3.2, find a 95% CI for the mean. Assume normality with unknown variance.

Frequentist confidence interval is $3.19 \pm t_{0.025,23} \left(\frac{0.2903}{\sqrt{24}} \right)$ which produces [3.00, 3.31].

Using the WinBUGS code:

```
model{
for (i in 1:n)
{#model
#df=n-1
x[i]~dt(mu, tau,23) #Syntax for t disn is dt(mu, tau, df). Each x is t distributed with randomly generated mu and tau and 23 df.
}
#priors
mu~dunif(0,10)
#tau~dgamma(0.01,0.01) Using either a uniform or a gamma distribution with small parameters produces similar results.
tau~dunif(0,24)
}
#Data
list( x=c(3.4,3.5,3,3.4,3.3,3.5,3.3,3,3.5,3,2.7,3.1,3.8,3.1,3.3,3.1,3.1,2.9,2.5,3.1,3,3.6,3.2,3.2), n=24)
#Initial values
list(mu=3, tau=10)
```

and calculating the mean, the Bayesian credible interval for the mean is [3.07, 3.31]. See Table 6. Students will ask the question that what will happen if you assume the variance is unknown in Example 6. Instructors should let them answer their question by themselves.

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|----|-------|---------|----------|----------|--------|-----------|-------|--------|
| mu | 3.193 | 0.06022 | 6.791E-4 | 3.073 | 3.193 | 3.311 | 10001 | 10000 |

Table 6.

Example 8: Find a 95% confidence interval for the variance of the data 16, 19, 22, 23, 18, 17, 18, 16, 11, 8, 8, 7, 9, 10, 12, 18, 19, 16, 12, 10, 11, 9, 9, 7, 6, 9, 10, 12, 15, 15. Kokoska, S (2015)

Frequentist 95% confidence interval is [14.257, 40.622].

Using the WinBUGS code:

```
model{
for (i in 1:30)
{#model
x[i]~dt(mu, tau,29)
}
#priors
mu~dunif(0,24)
#tau~dgamma(0.01,0.01)
tau~dunif(0,100)
variance<-1/tau
}
#Data
list( x=c(16,19,22,23,18,17,18,16,11,8,8,7,9,10,12,18,19,16,12,10,11,9,9,7,6,9,10,12,15,15))
#Initial values
list(mu=15, tau=50)
```

and calculating the mean the Bayesian credible interval for the variance is (12.67, 36.15). See Table 7. At this stage student will be struggling with the difference between confidence intervals and credible intervals. Actuarial science majors will be delighted to learn about the credible intervals. I usually ask students to read about it and prepare for a discussion in class. A two page paper before the discussion is also another approach if writing is part of the course.

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|----------|---------|---------|----------|----------|---------|-----------|-------|--------|
| tau | 0.04961 | 0.01311 | 1.535E-4 | 0.02767 | 0.04851 | 0.07893 | 10001 | 10000 |
| variance | 21.65 | 6.069 | 0.06796 | 12.67 | 20.62 | 36.15 | 10001 | 10000 |

Table 7.

Example 9: James H. Albert (1993) gives an example of Bayesian prediction of Kevin Mitchell's home run production for 1989 using the 1986-1988 data. Kevin Mitchell had 53 homeruns from 1297 at-bats from 1986-1988 producing an average home run rate of 0.041. Assuming the home run rate p is between 0 and 0.1 about 95% of the probability to be between 0.01 and 0.1 one can model the p by a Beta distribution with parameters a and b . Since a and b are unknown, the mean of the Beta distribution can

be used to get some idea of the values of a and b . Equation $\frac{a}{a+b} = \frac{0.10-0.01}{2}$ produces $b \approx 21a$. By trial and error and using WinBUGS code:

```
model{
x~dbeta(4,84) #x~dbeta(a, 21a)
}
```

and calculating the summary statistics, $a = 4$ produces a desired Beta distribution. See Table 8.

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|---|--------|---------|----------|----------|--------|-----------|-------|--------|
| x | 0.0455 | 0.02213 | 2.296E-4 | 0.01304 | 0.042 | 0.0979 | 1 | 10000 |

Table 8.

James Albert (1993) predicted the homerun rate for 50 future games assuming that Kevin Mitchell played 4 home runs in 44 at bats during the first 10 games of the 1989 season.

Using the WinBUGS code:

```
model{
y<-4
p~dbeta(4,84)
y~dbin(p,44)
ypred~dbin(p,50)
pprior~dbeta(4,84)
yprior~dbin(pprior,50)
#####OR
pnew~dbeta(8,124)
ynew~dbin(pnew,50)
z<-1/pnew
}
```

and calculating the summary statistics and graphing the prior and posterior probabilities one can reproduce the predictions of James Albert (1993) for 50 future games using WinBUGS. James Albert used MINITAB macros for his simulations. In this paper variable *yprior* does not consider the assumed 4 homeruns during first 10 games. Variables *ypred* and *ynew* are supposed to predict the distribution of home runs for the next 50 at-bats. Variable *z* estimates the number of times at bat to get one home run. See Table 9 and Figure 4.

| | mean | sd | MC_error | val2.5pc | median | val97.5pc | start | sample |
|--------|---------|---------|----------|----------|---------|-----------|-------|--------|
| p | 0.0609 | 0.02081 | 2.211E-4 | 0.02695 | 0.05881 | 0.108 | 10001 | 10000 |
| pnew | 0.06045 | 0.02048 | 1.748E-4 | 0.02686 | 0.05826 | 0.1066 | 10001 | 10000 |
| pprior | 0.0451 | 0.02197 | 2.305E-4 | 0.0123 | 0.04179 | 0.09677 | 10001 | 10000 |
| ynew | 3.029 | 1.956 | 0.01977 | 0.0 | 3.0 | 7.0 | 10001 | 10000 |
| ypred | 3.04 | 1.971 | 0.0194 | 0.0 | 3.0 | 8.0 | 10001 | 10000 |
| yprior | 2.252 | 1.822 | 0.01663 | 0.0 | 2.0 | 7.0 | 10001 | 10000 |
| z | 18.71 | 7.345 | 0.06918 | 9.383 | 17.16 | 37.3 | 10001 | 10000 |

Table 9.

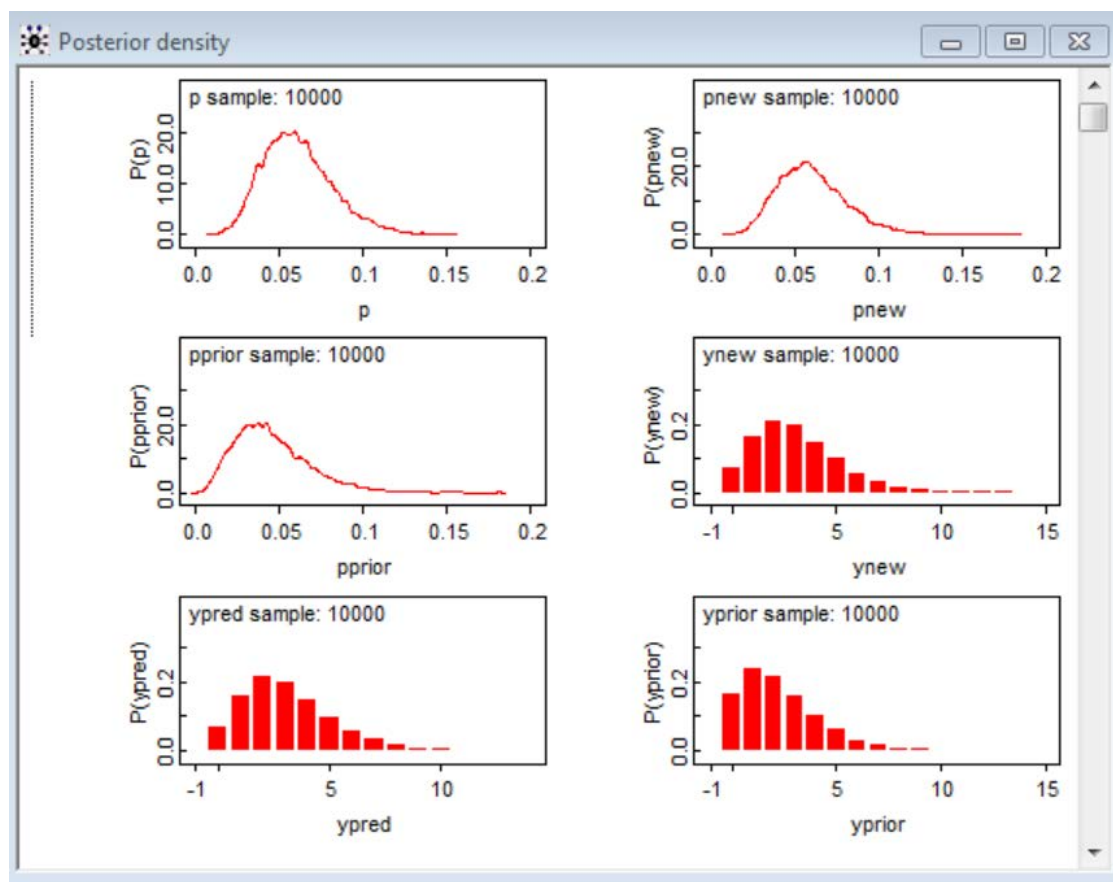


Figure 3.

Our results are comparable to James Albert's results but writing and running WinBUGS code did not take much time.

4. Conclusion

Students had difficulties with selecting the prior and switching between the familiar frequentist methods and using subjective Bayesian priors. Slowly they understood the difference between the non-informative prior and a subjective prior. They enjoyed the computations compared to analytical approaches. Students enjoyed the Monte Carlo approach to estimate integrals such as the area inside the unit circle and especially estimating the mean of a probability distribution. Some even commented that they understood calculus better after these exercises. Monte Carlo approach to explaining the p-value of a hypothesis test was another favorite example among the students. Most students appreciated the value of learning WinBUGS. As far as the workload for the instructor, some of these examples were done for the graduate students and some by the graduate students, the pleasure of teaching new things was more than the little burden of getting ready for classes. Bayesian software and computer technology had improved so much these days one can easily add Bayesian methods in to even general education elementary statistics course.

References

Albert, J. H. (1993) "Teaching Bayesian Statistics Using Sampling Methods and MINITAB." *The American Statistician*, Vol 47, No. 3, pp 182-191.

Berry, D. A. (1997) "Teaching Elementary Bayesian Statistics with Real Applications in Science." *The American Statistician*, Vol. 51, No. 3, pp 41-245.

Kokoska, S (2015) *Introductory Statistics: A Problem-Solving Approach*, Second ed. Freeman, New York.

Lunn et. al. (2013) *The BUGS Book*, CRC Press, Boca Raton, FL.

Rice, J. A. (2009) *Mathematical Statistics and Data Analysis*, Duxbury, 2009.