ST440/540 Applied Bayesian Analysis
Lab activity for 2/10/2025

Announcements

- No quiz Friday, but still videos to watch on Wednesday.
- Exam due next Monday, 1/17. Note, the exam could make use of material covered on Wednesday's video including the code in this example and at the end of this document.

A. STUDENT QUESTIONS

(1) How important are reference priors? Do we need to have a good understanding of which to use, or will we be given the a certain prior?

> I only ever use Jeffries' priors (and I use these rarely) and I only know the ones in the slides off the top of my head. In another other cases I would have to look them up or derive them.

(2) I would like to review the theory and application of selecting uninformative priors. Right now, I feel like I can only use what we learned in class to specify an uninformative prior, but generally, I would not know how to do this.

> I'd say the theory for uninformative priors is the Jeffries priors. In other cases, it's more ad hoc. Usually we just use Gamma(0.01,0.01) or Normal(0,100) priors since these have large variance (of course, "large" is problem specific).

(3) Is a zero-information improper prior (i.e. a Gamma(0, 0) distribution) considered an objective prior? Are any checks required other than evaluating that the posterior integrates to 1 over its support?

> An "objective prior" is one that is derived using a generally procedure like the Jeffreys' priors. So if a Gamma(0,0) turns out the be the Jeffreys' prior for a parameter, than yes. But if you just pick this prior without going through some procedure, then no.

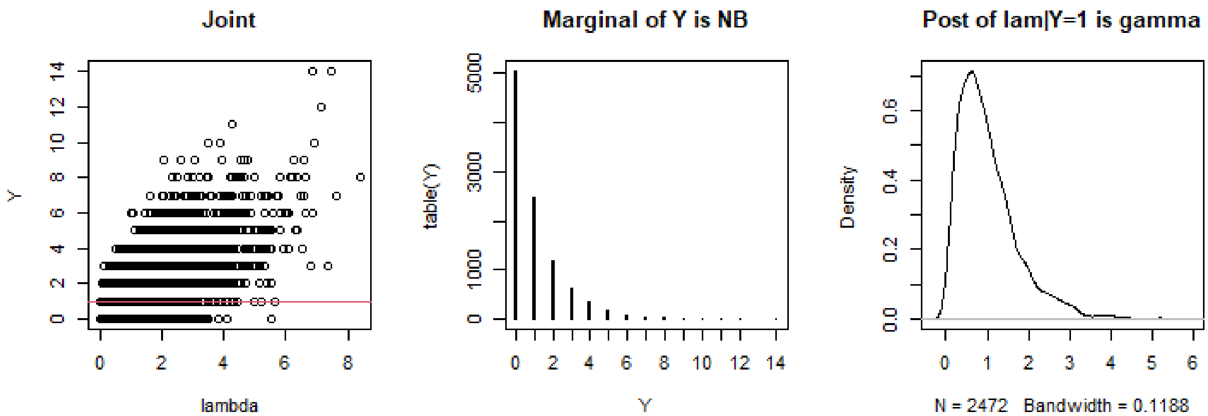
(4) What approach are you looking for when asking questions like "how sensitive are results to prior"? Should we just compare to a few other priors or is there a more systematic approach?

> Yes, exactly. There is a good example of this in the notes (near the end) from the first class where the model was fit for a few beta priors and a table summarized the posterior for each prior.

(5) Wikipedia's page on conjugate priors says that a poisson likelihood and gamma prior results in a negative binomial posterior, but other sources say this combination results in a gamma distribution for the posterior. Are both true?

If Y|lambda ~ Poisson(lambda) and lambda ~ Gamma then the marginal distribution of Y averaging over lambda is negative binomial, but the posterior of lambda|Y is gamma. So

```
S      <- 10000
lambda <- rgamma(S,1,1)
Y      <- rpois(S,lambda)

plot(lambda,Y,main="Joint");abline(1,0,col=2)
plot(table(Y),main="Marginal of Y is NB")
plot(density(lambda[Y==1]),main="Post of lam|Y=1 is gamma")
```



(6) Can you explain again what fisher's information is? I am having a hard time grasping the concept.

For a one-parameter case, the Information is the second derivative of the log likelihood, so

$$\frac{d^2}{d\theta^2}\log[f(Y|\theta)].$$

If there are p parameters then it is a p x p matrix with all the cross derivatives. This matrix is super important in frequentist statistics because it's from this that most standard errors are computed. It only shows up sporadically in a Bayesian analysis.

(7) This may be a bit of a random question/connection, but in your experience what kind of research areas typically utilize Bayesian analysis? I feel like it overlaps a lot with Game Theory.

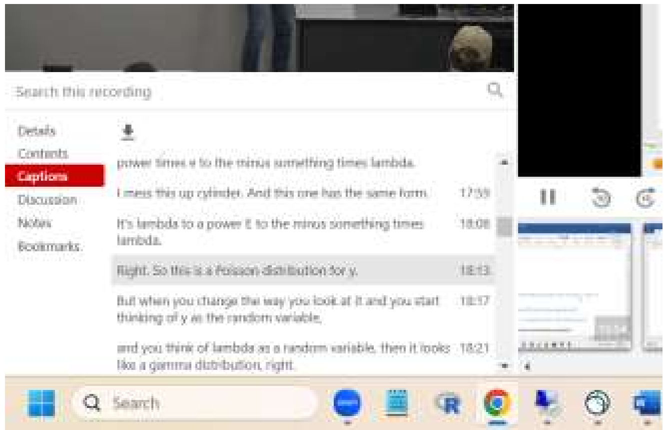
Bayesian methods are now used in a lot of fields. I see it most in ecology and environmental statistics, but it's really used everywhere now.

(8) Can you give any information on whether there is any relation between a prior's entropy and the distance between it and the posterior model? If so, how would this influence one's choice of a prior or likelihood?

I have never used entropy in an analysis so it is clearly not critical. I just added one slide about maximum entropy priors because it's a neat idea. I naively think of entropy as uncertainty/variance, so if a prior's entropy/variance is large it will have little effect on the posterior and vice versa.

(9) This isn't about this specific lecture, but it is possible to auto-generate subtitles for the lectures? I can't seem to toggle subtitles for the last few lectures (though I was able to the first few weeks).

For me, I hit “captions” in the bottom left and it come up. I imagine it varies a bit by machine.



(10) I noticed HW solutions aren't posted, so I wondered how the posterior pdf of rho would be determined for Chapter 1, Problem 12 by hand?

The homework solutions are given in the lab sheet, i.e., Section B below.

(11) Some questions on the quizzes seem to be coming from material that is not related to the assigned lectures. Are we expected to be working ahead to prepare?

Yeah, I mess this one up and copied and pasted a stray question about Metropolis. Sorry about this.

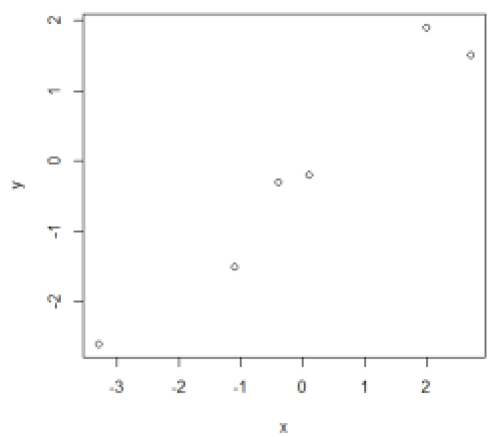
B. QUIZ AND HOMEWORK SOLUTIONS

Q3: Explain why the Jeffreys prior is a reasonable default prior.

It is the only prior that is invariant to transformations.

Chapter 1, problem 12: The data clearly show a strong positive correlation

```
x <- c(-3.3, 0.1, -1.1, 2.7, 2.0, -0.4)
y <- c(-2.6, -0.2, -1.5, 1.5, 1.9, -0.3)
plot(x,y)
```



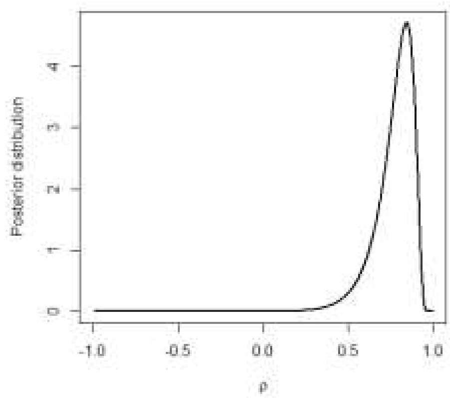
The bivariate normal PDF from Equation (1.25) on Page 23 with means equal 0 and variance equal 1 is

```
binorm <- function(x,y,rho){
  num  <- (x^2-2*rho*x*y+y^2)/(1-rho^2)
  like <- exp(-0.5*num)/sqrt(1-rho^2)
return(like)}
```

You could also use the `dmvnorm` function in the `mvtnorm` package. This gives the PDF of one (x,y) pair. There are six of these and they are independent, so the likelihood is the product of six PDFs. This is plotted on a grid below

```
nr   <- 1000
rho  <- seq(-0.999,0.999,length=nr)
post <- rep(1,nr)

for(r in 1:nr){
  for(i in 1:6){ # Likelihood
    post[r] <- post[r]*binorm(x[i],y[i],rho[r])
```

```
  }

  # Prior
  post[r] <- post[r]*dunif(rho[r],-1,1)
}

marg <- sum(post)*(rho[2]-rho[1])
post <- post/marg # Normalize the posterior

plot(rho,post,type="l",lwd=2,xlab=expression(rho),
                              ylab="Posterior distribution")
```



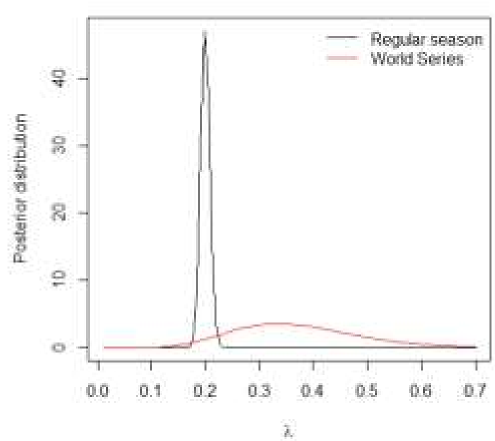
The posterior concentrates around rho=0.95.

Chapter 2, problem 2: The data are

```
Y_reg <- 563
N_reg <- 2820
Y_ws  <- 10
N_ws  <- 27
```

Since Y is a count, we assume Y|lambda ~ Poisson(N*lambda) for both regular season and world series. The conjugate prior for a Poisson rate is lambda ~ Gamma(a,b). Then the posterior is lambda|Y ~ Gamma(Y+a,N+b). To make an uninformative prior we select a=b=0.1. The code below applies this method separately for regular season and World Series games, and uses Monte Carlo sampling to compute the posterior probability that the Poisson rate is larger for the World Series than regular season.

```
lambda <- seq(0.01,0.7,length=100)
a <- b <- 0.1
plot(lambda,dgamma(lambda,Y_reg+a,N_reg+b),type="l",
                   xlab=expression(lambda),
                   ylab="Posterior distribution")
lines(lambda,dgamma(lambda,Y_ws+a,N_ws+b),col=2)
legend("topright",c("Regular season","World
Series"),lty=1,col=1:2,bty="n")
```




```
> lambda_ws  <- rgamma(100000,Y_ws+a,N_ws+b)
> lambda_reg <- rgamma(100000,Y_reg+a,N_reg+b)
> mean(lambda_ws>lambda_reg)
[1] 0.95297
```

The posterior probability that Mr. October has a higher home run rate in the World Series is 0.95, so fairly strong evidence to support his reputation.

B. DISCUSSION QUESTIONS

(1) Assume that Y is normal with mean zero and precision (inverse variance) τ. Assuming τ has a Gamma(a,b) prior, derive its posterior.

Hint: The PDF of Y|τ and τ are $f(y|\tau) = \frac{\tau^{1/2}}{\sqrt{2\pi}} e^{-\tau\frac{y^2}{2}}$ and $\pi(\tau) = \frac{b^a}{\Gamma(a)} \tau^{a-1} e^{-b\tau}$.

$$p(\tau|y) \propto f(y|\tau)\pi(\tau) \propto [\tau^{1/2} e^{-\tau\frac{y^2}{2}}][\tau^{a-1} e^{-b\tau}] \propto \tau^{A-1} e^{-B\tau}$$

where A=1/2+a and B = y^2/2+b, therefore τ|Y~Gamma(A,B).

(2) In a study of energy efficiency, n=100 buildings were equipped with new sensors that detect if a room is empty and adjust the room temperature accordingly. For building i=1,…,n, let Y_{0i} be the energy usage per day in the year before the equipment was installed, Y_{1i} be the energy usage per day after installation and $Y_i = Y_{1i} - Y_{0i}$. The objective is to test whether the new equipment reduces the average energy cost.

(a) Assuming all variance parameters are known, define a likelihood and objective Bayesian prior.

Likelihood: $Y_1,\ldots,Y_n|\mu \sim \text{Normal}(\mu,\sigma^2)$ independent for i = 1,…,n (we know σ)

Prior: π(μ) = 1 for all μ in (-∞,∞).

(b) What is the posterior for the model in (a)?

$\mu|Y_1,\ldots,Y_n \sim \text{Normal}(\bar{Y},\sigma^2/n)$ where $\bar{Y}$ is the sample mean of $Y_1,\ldots,Y_n$

(c) How would you summarize the posterior?

Compute the posterior probability that the mean of the difference is positive, i.e., $\text{Prob}(\mu > 0|Y_1,\ldots,Y_n)$ and if this probability is higher than 0.95 we conclude the mean is likely non-zero.

(d) How is this different than a frequentist analysis?

A frequentist analysis would do a one-sample z-test and decide to accept or reject based on the p-value, whereas a Bayesian analysis you get the post probability of each hypothesis. It turns out that in this simple case, the Bayesian estimate (posterior mean $\bar{Y}$) is the same as the frequentist estimate, the 95% credible set is exactly the 95% confidence interval and the frequentist z-test gives the same decision as the Bayesian test.

(e) Name a source of bias that could lead to concluding the new equipment is effective even if it's true effect is zero (other than random chance). Give a tweak to the design of the experiment that could avoid this bias.

If all the treatments were applied at the beginning of 2021 and 2021 happened to be cooler than 2020 you could incorrectly conclude the equipment if effective. One approach to avoiding this bias is to randomize some buildings to receive the treatment and some to serve as controls, then compare the means between the two groups.

(3) If Y|p~Binomial(n,p) we saw in the video that the Jeffreys' prior (JP) for p is p~Beta(1/2,1/2). Now say our primary interest was to analyze the odds q = p/(1-p). The last page shows the JP for q is

$$\pi(q) \propto \frac{1}{\sqrt{q}(q+1)}$$

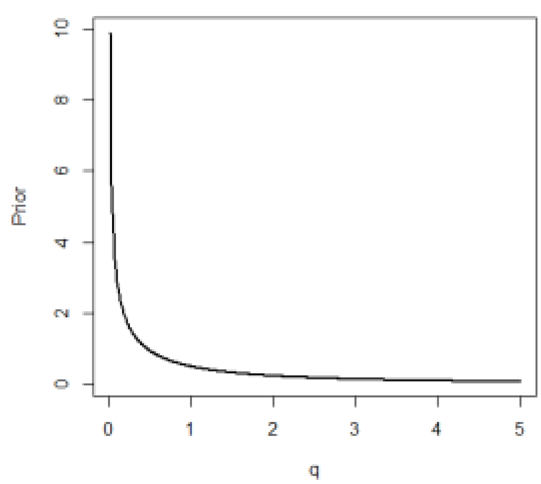
(a) Describe the JP for q. Is it proper? Which values are favored?

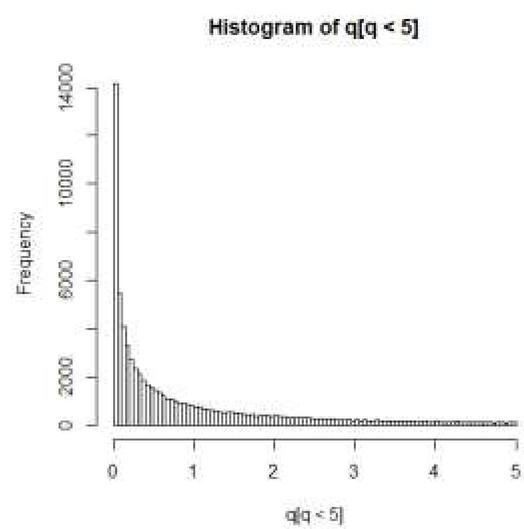
It is a proper PDF since $\int_0^\infty \pi(q)dq < \infty$. This follows from the calculus result that $\int_0^\infty \frac{1}{q^k}dq < \infty$ if k>1. The PDF has mode at zero and decays with q.

(b) Explain the consequences of the plots below.

This shows that if we start with the JP for p and convert to q (plot on the right) we get the same thing as if we directly derive the JP for q (plot on the left). This demonstrates that the JP is invariant to reparameterization.

```
> q <- seq(0.01,5,0.01)
> plot(q,1/(sqrt(q)*(1+q)),type="l",xlab=expression(q),ylab="Prior",lwd=2)
> p <- rbeta(100000,0.5,0.5)
> q <- p/(1-p)
> summary(q)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.000e+00 0.000e+00 1.000e+00 3.019e+04 6.000e+00 1.047e+09
> hist(q[q<5],breaks=100)
```





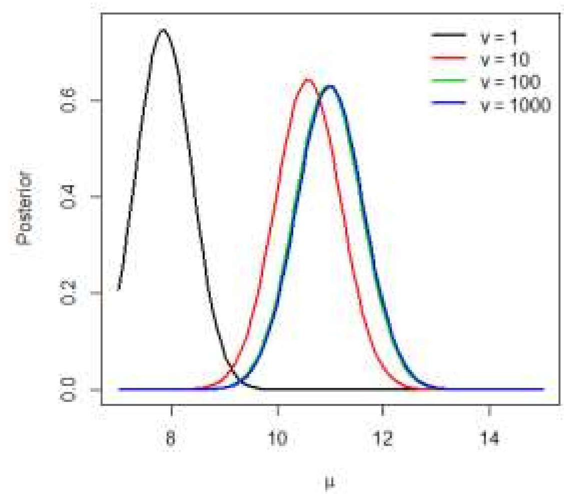
(4) Say $Y_1,\ldots,Y_n$ are Gaussian with mean μ and variance 5. The sample size is n=10 and the sample mean is $\bar{Y} = \sum_{i=1}^{n} Y_i/n = 11$. Assuming prior μ ~ Normal(0,v), the plots below show the posterior for several values of v.

(a) Would you say the posterior is sensitive to the prior?

To a point yes, but for v>10 the posterior changes only a bit with v.

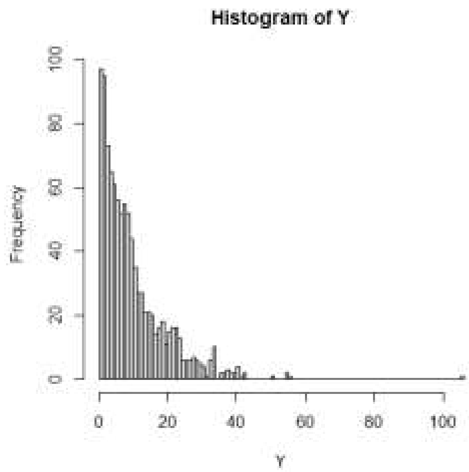
(b) How would you present the result of the analysis?

I would present the results for v=100 and say the results are similar for v = 10 and 1000.




```
n      <- 10
sig2   <- 4
ybar   <- 11
m      <- c(0,0,0,0)
v      <- c(1,10,100,1000)
y      <- ybar+4*seq(-1,1,length=100)
plot(NA,xlim=range(y),ylim=c(0,0.75),xlab=expression(mu),ylab="Posterior")
for(j in 1:4){
   vvv <- 1/(n/sig2 + 1/v[j])
   mmm <- n*ybar/sig2 + m[j]/v[j]
   lines(y,dnorm(y,vvv*mmm,sqrt(vvv)),type="l",lwd=2,col=j)
}
legend("topright",paste("v =",v),lwd=2,col=1:4,bty="n")
```

(5) Our analysis of the vaccine trial data in the last lab was actually a simplification of the analysis used in the paper. They use survival analysis methods. Let the response Y be the time (measured in weeks) from enrolling in the study until the time of testing positive for COVID. The (totally made up) data are



(a) What are some pros and cons of analyzing time until infection versus a binary indicator that a subject is infected in the first three months?

Time until infection uses all the data, but requires picking a family of distributions for the response and waiting until all patients have been infected to conduct the analysis (actually, you can analyze the data before everyone has been infected using *censored* survival analysis methods).

(b) Propose a family of distributions for the data plotted above.

The data look like an Exponential(θ) distribution. That is, $Y_i|\theta \sim \text{Gamma}(1, \theta)$ with PDF $f(y_i|\theta) = \theta \exp(-\theta y_i)$. Combining the n observations $y_1,\ldots,y_n$ gives likelihood function

$$f(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} [\theta e^{-\theta y_i}] = \theta^n e^{-\theta \sum_{i=1}^{n} y_i}$$

For notational simplicity, let $Y = \sum_{i=1}^{n} y_i$ for the rest of the problem.

(c) Derive a conjugate prior and the corresponding posterior for the parameter(s) in (b).

The likelihood is the kernel of a gamma distribution, so the conjugate prior is θ ~ Gamma(a,b). The posterior is then

$$p(y|\theta) \propto \left[\theta^n e^{-\theta Y}\right]\left[\theta^{a-1} e^{-\theta b}\right] \propto \theta^{n+a-1} e^{-\theta[b+Y]}$$

so the posterior is $\theta|y \sim Gamma(n + a, \sum_{i=1}^{n} y_i + b)$.

(d) Derive a Jeffries prior and corresponding posterior for the parameter(s) in (b).

The log likelihood is (ignoring constants that don't depend on θ)

$$\log[f(y|\theta)] = n log(\theta) - \theta Y$$

The first derivative with respect to θ is n/θ - Y. The second derivative is $-n/\theta^2$. Therefore, the Fisher's info is n/θ^2 and the prior is the square root, $\pi(\theta) \propto 1/\theta = \theta^{-1}$. The posterior is then

$$p(y|\theta) \propto \left[\theta^n e^{-\theta Y}\right]\left[\theta^{-1}\right] \propto \theta^{n-1}e^{-\theta Y}$$

so the posterior is $\theta|y \sim Gamma(n, \sum_{i=1}^{n} y_i)$. This is equivalent to a conjugate prior with a=b=0 and is proper for any n>0.

(e) Now say the data from treatment and placebo groups are analyzed separately using the conjugate prior, how would you summarize these posteriors to determine if the survival distribution is affected by the treatment?

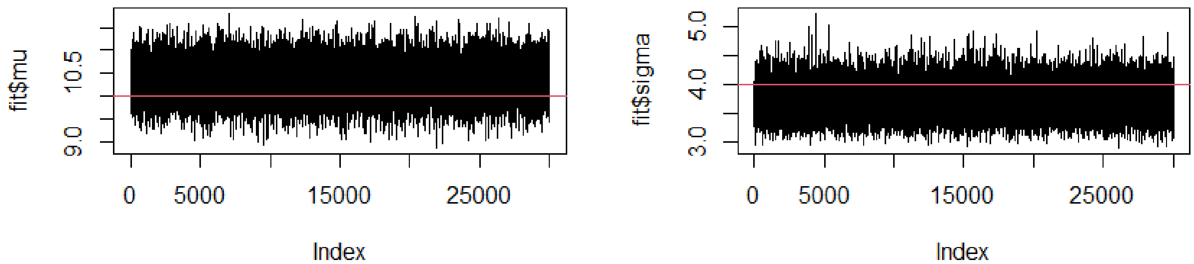
I make a huge sample of θ from the posterior for the treatments and separately for the controls, and then compute the proportion of the samples with larger θ for treatment than controls and use this to determine which group has longer survival time.

Potentially useful for Exam 1

A function to perform Gibbs sampling for a one-sample t-test, described here

https://www4.stat.ncsu.edu/~bjreich/BSMdata/gibbs_ttest.html

```
normal_Gibbs <- function(Y, n.iters = 30000,
                         m=0.1, a=0.1, b=0.1){ # define the priors

  n             <- length(Y)

  # Initial values
  mu            <- mean(Y)
  s2            <- var(Y)

  keep.mu       <- rep(0,n.iters)
  keep.sigma    <- rep(0,n.iters)
  keep.mu[1]    <- mu
  keep.sigma[1] <- sqrt(s2)

  for(iter in 2:n.iters){

    # sample mu|s2,Y

     MN  <- sum(Y)/(n+m)
     VR  <- s2/(n+m)
     mu  <- rnorm(1,MN,sqrt(VR))

    # sample s2|mu,Y

     A  <- a + n/2
     B  <- b + sum((Y-mu)^2)/2
     s2 <- 1/rgamma(1,A,B)

    # keep track of the results
     keep.mu[iter]    <- mu
     keep.sigma[iter] <- sqrt(s2)
  }

 out <- list(mu=keep.mu,sigma=keep.sigma)
return(out)}

# Apply the function to simulated data to verify it works
set.seed(919)
Y   <- rnorm(100,10,4)
fit <- normal_Gibbs(Y)
plot(fit$mu,type="l"); abline(10,0,col=2)
plot(fit$sigma,type="l"); abline(4,0,col=2)
```



fit$mu
9.0
10.5
0
5000
15000
25000
Index
fit$sigma
3.0
4.0
5.0
0
5000
15000
25000
Index