ST440/540 Applied Bayesian Analysis

Lab activity for 2/12/2024

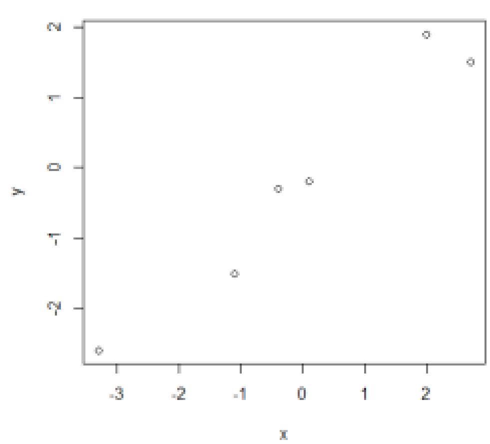
A. QUIZ AND HOMEWORK SOLUTIONS

Q3: We gather n observations and fit the model $Y_i \sim \text{Normal}(\mu,\sigma^2)$. Assuming σ is known but μ is not known and we select an uninformative prior. Our goal is to make a prediction for a new observation.

(a) Define parametric uncertainty: posterior uncertainty about the parameter μ
(b) Define sample/error uncertainty: Uncertainty about Y given μ, i.e., normal error with variance σ^2
(c) Which of these two types of error dominates for large n? Error uncertainty because the posterior variance of μ will decrease with n

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

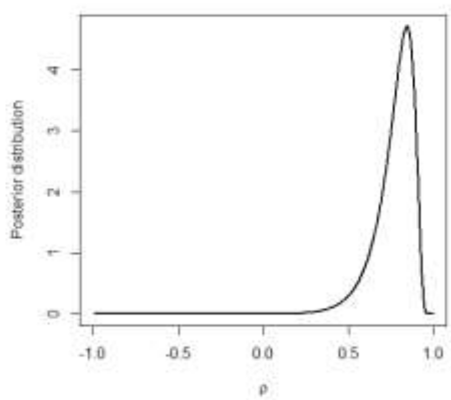
```
nr   <- 1000
rho  <- seq(-0.999,0.999,length=nr)
post <- rep(1,nr)

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")
```



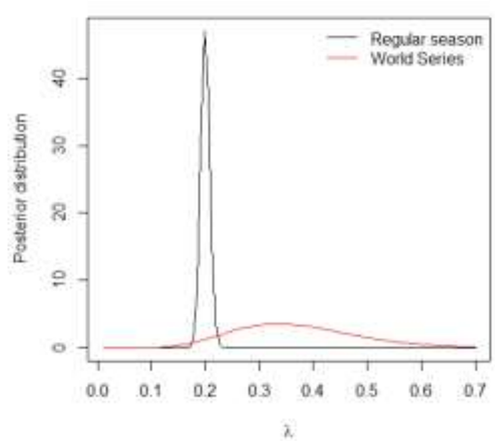
The posterior concentrates around rho=0.95.

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: $Y_1,\dots,Y_n|\mu \sim \text{Normal}(\mu,\sigma^2)$ 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)?

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

(c) How would you summarize the posterior?

Compute the posterior probability that the mean of the difference is positive, i.e., Prob(μ >0|$Y_1,\dots,Y_n$) 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.

(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

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

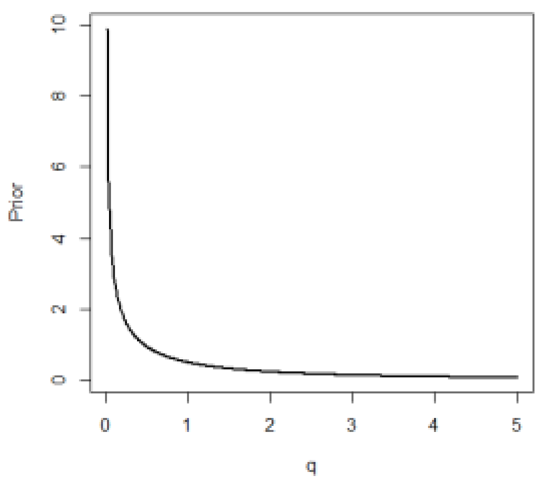
(a) Describe the JP for q. Is it proper? Which values are favored?

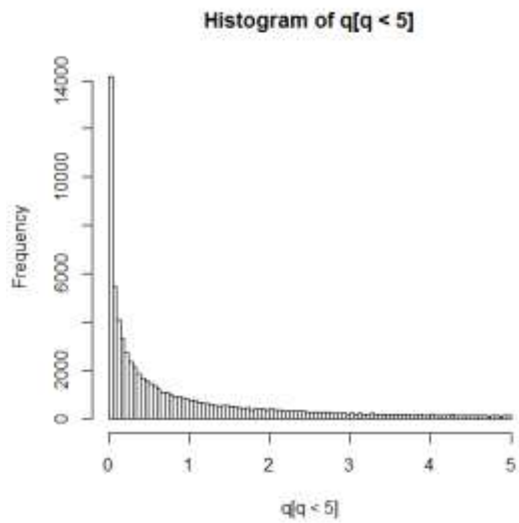
It is a proper PDF since $\int_0^\infty \pi(q)dq < \infty$. This follows from the calculus result that $\int_0^\infty \frac{1}{q^k}dq < \infty$ if k>1. The PDF has mode at zero and decays with q.

(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)
```





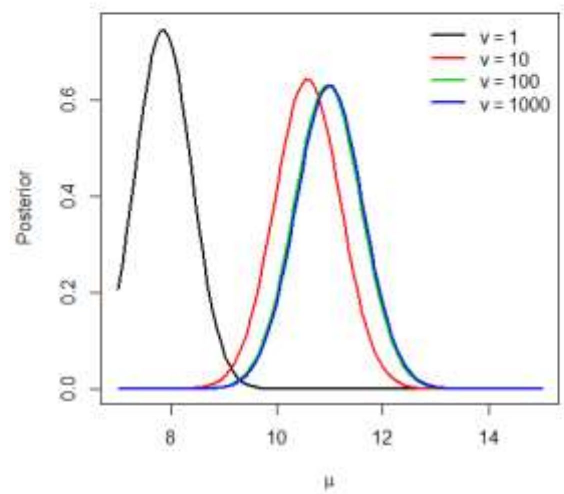
(4) Say $Y_1,\ldots,Y_n$ are Gaussian with mean μ and variance 5. The sample size is n=10 and the sample mean is $\bar{Y} = \sum_{i=1}^{n} Y_i/n = 11$. Assuming prior μ ~ Normal(0,v), the plots below show the posterior for several values of v.

(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:

https://en.wikipedia.org/wiki/Survival_analysis

The details are given on page 15 of

https://www.nejm.org/doi/suppl/10.1056/NEJMoa2035389/suppl_file/nejmoa2035389_appendix.pdf

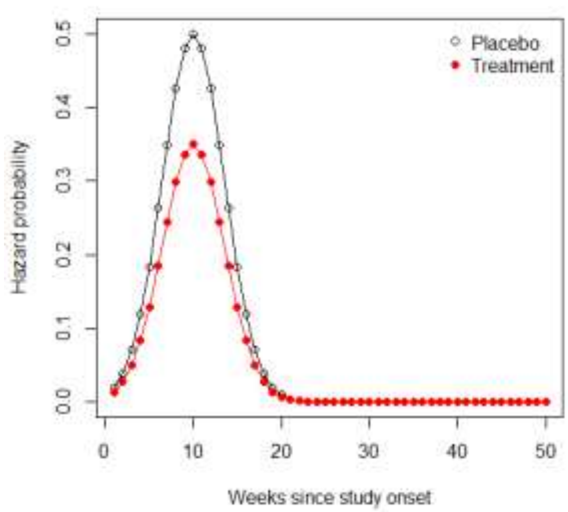
In short, rather the response for a patient being the binary indicator that they got infected, the response is the time until they became infected. Patients that never get infected are called *censored* because we don't know their exact response at the time of conducting the analysis.

They define efficacy via the hazard ratio. The hazard function at time t days after vaccination is defined as the probability of getting infected on day t given you have not been infected prior to day t. They assume that the hazard functions of the two treatment groups are proportional to each other, and the ratio of the hazard functions defines efficacy. The (discrete time) proportional hazards model is

P(infected on day t | not infected prior to day t and placebo) = lambda(t)
P(infected on day t | not infected prior to day t and vaccine) = lambda(t)*beta

and beta is the hazard ratio and the parameter of interest is E=1-beta. Here is hypothetical example

```
> t      <- 1:50
> lambda <- 0.5*exp(-((t-10)/5)^2)
> beta   <- 0.7
> plot(t,lambda,col=1,type="l",
+      xlab="Weeks since study onset",
+      ylab="Hazard probability")
> lines(t,lambda*beta,col=2)
> points(t,lambda,pch=1)
> points(t,lambda*beta,pch=19,col=2)
> legend("topright",c("Placebo","Treatment"),col=1:2,pch=c(1,19),bty="n")
```



Bringing this back to Bayes, say the time until infection is $Y|\lambda \sim$ Exponential(λ) with PDF and CDF

$$f(y|\lambda) = \lambda e^{-\lambda y} \text{ and } F(y|\lambda) = Prob(Y < y|\lambda) = 1 - e^{-\lambda y}$$

Now say that the trial ends in 10 weeks. If Y<10 then we observe the time of infection. But if Y>10 then we do not, and all we know is that the patient was still uninfected at the end of the study. Give the contribution to the likelihood for

(a) A patient with first infection at Y=5 weeks from vaccination

$$f(5|\lambda) = \lambda e^{-\lambda 5}$$

(b) A patient that is censored, i.e., not infected at the conclusion of the study

$$Prob(y > 10|\lambda) = 1 - F(y|\lambda) = e^{-\lambda 10}$$

(c) Three patients with Y_1=3, Y_2=7 and Y_3>10

$$f(3|\lambda)*f(7|\lambda)*[1 - F(10|\lambda)] = \lambda^2 e^{-\lambda(3+7+10)}$$

(d) Can you find a conjugate prior for λ?

The likelihood in (c) has the form of a gamma distribution for λ. If we pick prior $\lambda \sim$ Gamma(a,b) then the posterior is

$$p(\lambda|y) \propto f(3|\lambda)*f(7|\lambda)*[1 - F(10|\lambda)]\ \pi(\lambda) = [\lambda^2 e^{-\lambda(3+7+10)}][\lambda^{a-1} e^{-\lambda b}] \propto \lambda^{A-1} e^{-\lambda B}$$

where A = 2+a and B = 17+b, therefore $\lambda\ |y \sim$ Gamma(A,B).

Derivation for the Jeffreys' prior for the odds

The odds are $q = \frac{p}{1-p}$ so $p = \frac{q}{q+1}$ because

$$q(1-p) = p \Rightarrow q = p + pq \Rightarrow p = \frac{q}{1+q}$$

So the model written in terms of the odds is

$$Y|q \sim \text{Binomial}\left(n, \frac{q}{q+1}\right)$$

The log likelihood & its derivatives are

$$1-p = 1 - \frac{q}{q+1} = \frac{q+1-q}{q+1} = \frac{1}{q+1}$$

$$\ell(q) = \log\binom{n}{y} + y\log\left(\frac{q}{q+1}\right) + (n-y)\log\left(\frac{1}{q+1}\right)$$

$$= \log\binom{n}{y} + y\log(q) - y\log(q+1) - (n-y)\log(q+1)$$

$$= \log\binom{n}{y} + y\log(q) - n\log(q+1)$$

$$\ell'(q) = \frac{y}{q} - \frac{n}{q+1}$$

$$E(Y) = np = n\frac{q}{1+q}$$

$$\ell''(q) = -\frac{y}{q^2} + \frac{n}{(q+1)^2}$$

$$-E(\ell''(q)) = n\frac{q}{1+q}\frac{1}{q^2} - \frac{n}{(q+1)^2}$$

$$= \frac{n}{(1+q)q} - \frac{n}{(q+1)^2} = n\frac{1+q-q}{q(q+1)^2} = \frac{n}{q(q+1)^2}$$

So $\pi(q) \propto \sqrt{-E(\ell''(q))} \propto \dfrac{1}{q^{\frac{1}{2}}(q+1)}$