ST559 Homework 4

Nick Sun

May 4, 2020

Abstract

Problems 1, 4, 6, 10 from chapter 3 of *Bayesian Data Analysis*

1 Q1

Suppose data $(Y_1, \ldots, Y_n)$ follows a multinomial distribution with parameters $(\theta_1, \ldots, \theta_n)$. Also, suppose that $\theta = (\theta_1, \ldots, \theta_n)$ has a Dirchlet prior distribution. Let $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$.

1.1 part a.

Write the marginal posterior distribution for α.

answer It is known that the Dirichlet prior distribution is conjugate to the multinomial distribution so that the posterior is also Dirichlet. We can show this using some quick calculations.

$$\begin{aligned} p(\theta|data) &\propto p(data|\theta)\pi(\theta) \propto \theta_1^{y_1}\theta_2^{y_2}\ldots\theta_n^{y_n}\prod_{i=1}^{n}\theta_i^{a_i-1} \\ &\propto \theta_1^{y_1+a_1-1}\ldots\theta_n^{y_n+a_n-1} \end{aligned}$$

Therefore, the posterior distribution is Dirichlet$(y_1 + a_1, \ldots, y_n + a_n)$. This is very useful since we can use the property that the marginal distributions of a Dirchlet distribution are also Dirichlet. Therefore we get the following pdf for θ_1, θ_2:

$$p(\theta_1, \theta_2|data) \propto \theta_1^{y_1+a_1-1}\theta_2^{y_2+a_2-1}(1-\theta_1-\theta_2)^{\sum_{i=3}^{n}(y_i-a_i)-1}$$

If $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$, lets define $\beta = \theta_1 + \theta_2$. This will give us $\theta_1 = \alpha\beta$, $\theta_2 = (1-\alpha)\beta$, and $1-\theta_1-\theta_2 = (1-\beta)$. Now we just need to do a bivariate transformation to get the joint pdf for α, β.

The Jacobian for this transformation will be:

$$\begin{aligned} J &= \begin{pmatrix} \frac{\partial\alpha}{\partial\theta_1} & \frac{\partial\alpha}{\partial\theta_2} \\ \frac{\partial\beta}{\partial\theta_1} & \frac{\partial\beta}{\partial\theta_2} \end{pmatrix} \\ &= \frac{1}{\beta^2} \end{aligned}$$

Now plugging this in, we get the following joint posterior.

$$\begin{aligned} p(\alpha, \beta|data) &\propto (\alpha\beta)^{y_1+a_1-1}((1-\alpha)\beta)^{y_2+a_2-1}(1-\beta)^{\sum_{i=3}^{n} y_i-1} \\ &\propto \alpha^{y_1+a_1-1}(1-\alpha)^{y_2+a_2-1}\beta^{y_1+a_1+y2+a_2-2}(1-\beta)^{\sum_{i=3}^{n} y_i-1} \end{aligned}$$

Note that the joint posterior distribution is factorizable into distinct parts for α and β. Therefore, the posterior distribution for α is Beta$(y_1 + a_1, y_2 + a_2)$.

1.2 part b.

Show that this distribution is identical to the posterior distribution for α obtained by treating y_1 as an observation from the binomial distribution with probability α and samples size $y_1 + y_2$, ignoring the data $y_3, \ldots, y_n$.

answer In this problem, we assume that

$$Y_1 \sim \binom{y_1 + y_2}{y_1} \alpha^{y_1}(1-\alpha)^{y_2}$$

which is possible since α is a ratio and guaranteed to be between 0 and 1. If we use the fact that the prior for α should be a Beta distribution with a_1 and a_2 as the shape parameters, we get the following posterior.

$$\begin{aligned} p(\alpha|data) &\propto p(data|\alpha)\pi(\alpha) \propto \alpha^{y_1}(1-\alpha)^{y_2}\alpha^{a_1-1}(1-\alpha)^{a_2-1} \\ &\propto \alpha^{y_1+a_1-1}(1-\alpha)^{y_2+a_2-1} \end{aligned}$$

Therefore, the posterior is $p(\alpha|data) \propto$ Beta$(y_1 + a_1, y_2 + a_2)$ which is what we showed above!

2 Q4

An experiment was performed to estimate the effect of beta-blockers on mortality of cardiac patients. A group of patients were randomly assigned to treatment and control groups:

1. 647 patients received the control, 39 died.
2. 680 patients received the treatment, 22 died.

Assume that the outcomes are independent and binomially distributed, with probabilities of death p_0 and p_1 under the control and treatment respectively.

2.1 part a.

Set up a noninformative prior distribution for (p_0, p_1) and obtain posterior simulations.

answer We know that each group should follow a binomial distribution with p_0 for the control group and p_1 for the treatment group. Furthermore, we know that the groups are independent from each other and $p_0 \perp p_1$. A reasonable prior for each p is known to be Beta$(1/2, 1/2)$ and since they are independent, a reasonable joint prior distribution is $\pi(p_0, p_1) = \text{Beta}(p_0, 1/2, 1/2) * \text{Beta}(p_1, 1/2, 1/2)$.

To get the form of the posterior, we can plug in the data we obtained.

$$\begin{aligned} p(p_0, p_1|data) &\propto p(data|p_0, p_1)\pi(p_0, p_1) \\ &\propto p_0^{39}(1-p_0)^{635-1/2}p_1^{22-1/2}(1-p_1)^{658-1/2} \end{aligned}$$

Which again is the form of a joint Beta distribution.

We can use `R` to calculate some posterior simulations. Here, let's plot the posterior distribution for p_0 and p_1.

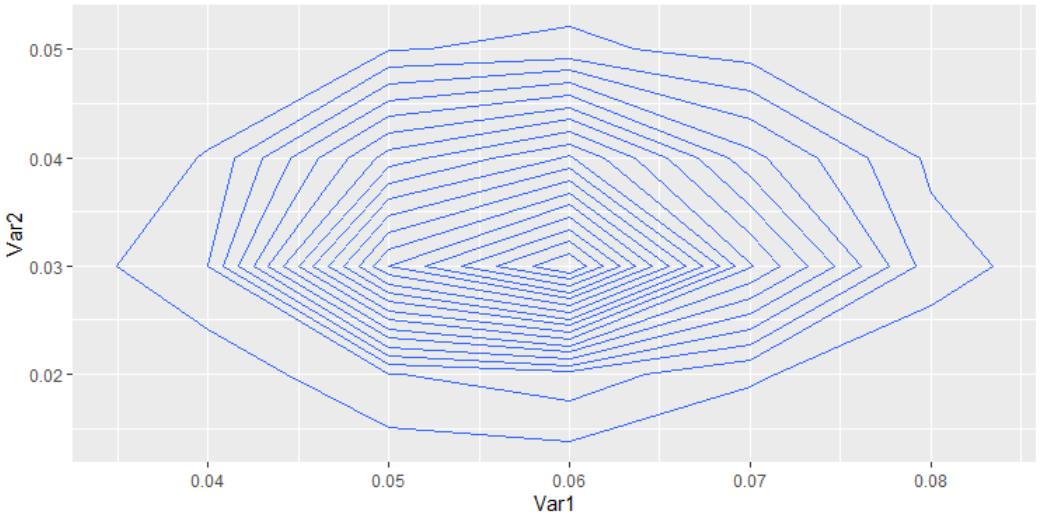
```
p0 <- seq(0, 1, .01)
p1 <- seq(0, 1, .01)
p0p1 <- expand.grid(p0, p1)

posterior <- function(p0, p1) {
    return((p0^(38.5))*((1-p0)^(634.5))*
  (p1^(21.5))*((1-p1)^(657.5)))
}

posts <- vector(mode = "numeric", length = nrow(p0p1))
for (i in 1:nrow(p0p1)) {
  posts[i] <- posterior(p0p1[i, "Var1"], p0p1[i, "Var2"])
}
p0p1$posts <- posts

ggplot(p0p1) +
geom_contour(mapping = aes(x = Var1,
                           y = Var2,
                           z = posts),
             bins = 20)
```

The resulting plot looks like this:



We can see that the posterior distribution is centered around $p_0 = .06$ and $p_1 = .03$. This is the equivalent to the MLEs $\hat{p}_0, \hat{p}_1$ that we would get on the control and treatment groups.

We could also generate some random observations from the posterior distribution using the following code.

```
p0.posts <- rbeta(1000, shape = 39.5, shape2 = 635.5)
p0.posts <- rbeta(1000, shape = 22.5, shape2 = 658.5)
```

2.2 part b.

Summarize the posterior distribution for the odds ratio $\frac{p_0}{1-p_0)}$ and $\frac{p_1}{1-p_1}$.

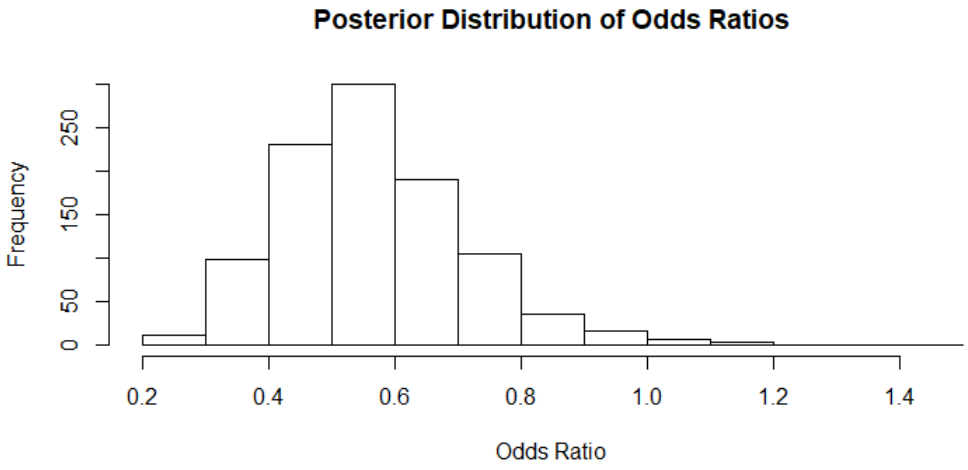
answer Again, we can use `R` to calculate this for us. Let's use the fact that we know the posterior distribution for both p_0 and p_1 with the data observed. We can use this to draw random observations from the posterior distribution and then calculate the odds ratio of these simulated posterior.

```
posterior_p0 <- rbeta(1000, 39.5, 635.5)
posterior_p1 <- rbeta(1000, 22.5, 658.5)
data.frame(
    p0 = posterior_p0,
  p1 = posterior_p1
) -> p0p1

odds_ratio <- function(p0, p1) {
  return((p1/(1-p1))/(p0/(1-p0)))
}

apply(p0p1, 1, function(row) odds_ratio(row[1], row[2])) ->
    posterior_odds_ratios
hist(posterior_odds_ratios)
summary(posterior_odds_ratios)
quantile(posterior_odds_ratios, c(.025, .975))
```

We get the following histogram for the posterior odds ratio.



A basic 5 number summary of this odds ratio is provided below.

Min	Q1	Median	Q3	Max
.2528	.4648	.5439	.6505	1.4044

with a mean of .5666. The 95% posterior interval is $(.3222, .9057)$ which does not contain 1. So we are fairly confident that the odds ratio is less than 1 and therefore the odds of death is lower in the treatment group than the control group.

2.3 part c.

Discuss the senstitivity of your inference to your choice of noninformative prior density.

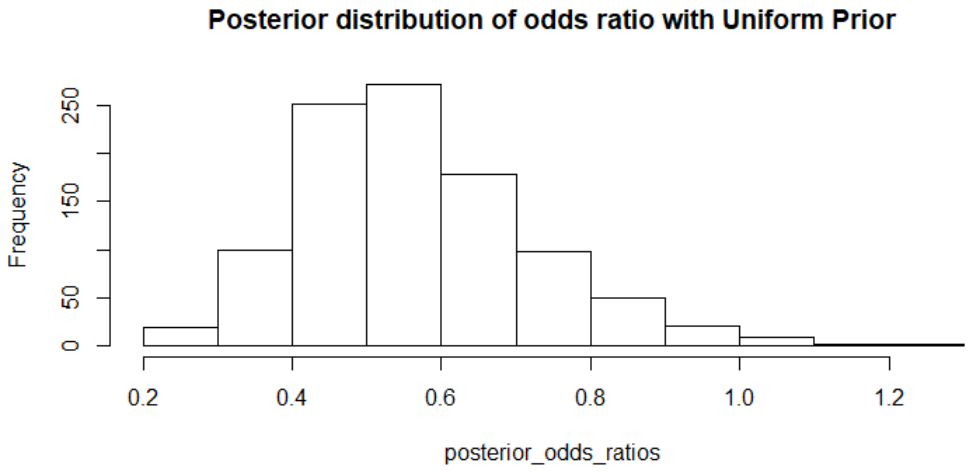
answer One way that we can go about solving this is to dry other priors and see how much the posterior changes. Since we used Jeffrey's prior in the parts above, let's try the uniform distribution which was the flat prior originally used by Laplace. The "nice thing" about the uniform distribution in this case is that it can be parameterized as a Beta(1, 1) distribution so we actually don't have to change our code that much.

```
posterior_p0 <- rbeta(1000, 40, 636)
posterior_p1 <- rbeta(1000, 23, 659)
data.frame(
    p0 = posterior_p0,
  p1 = posterior_p1
) -> p0p1

odds_ratio <- function(p0, p1) {
  return((p1/(1-p1))/(p0/(1-p0)))
}

apply(p0p1, 1, function(row) odds_ratio(row[1], row[2])) ->
    posterior_odds_ratios
hist(posterior_odds_ratios,
    main = "Posterior distribution of odds ratio with Uniform Prior")
summary(posterior_odds_ratios)
quantile(posterior_odds_ratios, c(.025, .975))
```

With a uniform prior, our histogram looks like this



and our 95% confidence interval for the posterior odds ratio is (.318, .923). While this isn't a huge deviation from the posterior we had with the Beta$(1/2, 1/2)$ prior, it is potentially meaningful.

After playing around with prior Beta distributions a bit and trying different shape parameters, I found that changing the shape parameters by 1 does have a noticeable impact on the posterior 95% interval and also seems to make the posterior distribution more skewed. Since this data directly deals with people's lives, I would say that this posterior probability is sensitive to the choice of prior and that we should probably stick to using a noninformative prior like Jeffrey's.

3 Q6

Consider data $Y_1, \ldots, Y_n$ modeled as independent $Bin(N, \theta)$ with both N and θ unknown. Defining a convenient family of prior distributions on (N, θ) is difficult, partly because of the discreteness of N.

A hierarchical approach based on assigning the parameters N a poisson distribution with unknown mean μ. To define a prior distributino on (θ, N), define $\lambda = \mu\theta$ and specifies a priro distribution on (λ, θ). The prior distribution is specified in term of λ rather than μ because it "seems easier to formulate prior information about λ, the unconditional expectation of the observations, than about μ, the mean of the unobserved quantity N".

3.1 part a.

A suggested noninformative prior distribution is $p(\lambda, \theta) \propto \lambda^{-1}$. What is the motivation for this noninformative distribution? Is the distribution improper? Transform to determine $p(N, \theta)$.

answer I'm not quite sure what the motivation for the prior distribution being proportion to λ^{-1} is, but it likely has to do with some assumptions about which parameters are uniform. For example, a reasonable assumption would be to have θ be distributed uniformly. Having the prior distribution proportional to λ^{-1} probably helps with that.

I do know that this prior distribution is improper since the integral $\int_0^\infty \frac{1}{\lambda} d\lambda$ diverges. To get $p(N, \theta)$, we can first find the joint probability of $p(N, \lambda, \theta)$.

$$p(N, \lambda, \theta) = p(N|\lambda, \theta)\pi(\lambda, \theta) = \frac{\left(\frac{\lambda}{\theta}\right)^N exp\left(-\frac{\lambda}{\theta}\right)}{N!\lambda}$$

Now integrating over λ, we get

$$\begin{aligned} p(N, \theta) &= \int_0^\infty \frac{\left(\frac{\lambda}{\theta}\right)^N exp\left(-\frac{\lambda}{\theta}\right)}{N!\lambda} d\lambda \\ &= \frac{1}{N!\theta^N} \int_0^\infty \lambda^{N-1} exp\left(-\frac{\lambda}{\theta}\right) d\lambda \\ &= \frac{\Gamma(N)\theta^N}{N!\theta^N} = \frac{1}{N} \end{aligned}$$

3.2 part b.

The Bayesian method is illustrated on counts of waterbuck obtained by remote photography on five separate days. The counts were $53, 57, 66, 67, 72$. Perform the Bayesian analysis on these data and display a scatterplot of posterior simulations (N, θ). What is the posterior probability that $N > 100$.

answer Let's derive the posterior distribution $p(N, \theta|data)$ first. We have five data points, assumed to have been drawn from some unknown binomial distribution.

$$\begin{aligned} p(N, \theta|data) = P(data|N, \theta)\pi(N, \theta) &= \prod_{i=1}^5 \binom{N}{y_i} \theta^{y_i}(1-\theta)^{N-y_i} \frac{1}{N} \\ &= \left(\prod_{i=1}^5 \binom{N}{y_i}\right) \theta^{\sum y_i}(1-\theta)^{\sum N - y_i} \frac{1}{N} \\ &\propto \theta^{\sum_{i=1}^5 y_i}(1-\theta)^{5N - \sum_{i=1}^5 y_i} \end{aligned}$$

Therefore, the posterior distribution is Beta$(\left(\sum_{i=1}^5 y_i\right) + 1, (5N - \sum y_i) + 1)$. We will also need to marginal posterior distribution for N.

$$\begin{aligned} p(N|data) &= \frac{1}{N} \prod_{i=1}^n \binom{N}{y_i} \int_0^1 \theta^{\sum y_i}(1-\theta)^{Nn - \sum y_i + 1} \\ &= \frac{1}{N} \prod_{i=1}^n \binom{N}{y_i} \text{Beta}\left(\sum y_i + 1, Nn - \sum y_i + 1\right) \end{aligned}$$

Note that we made use of the fact that the quantity underneath the integrand was the kernel of a Beta distribution. This posterior distribution is irregular and doesn't seem to fit any well known distribution.

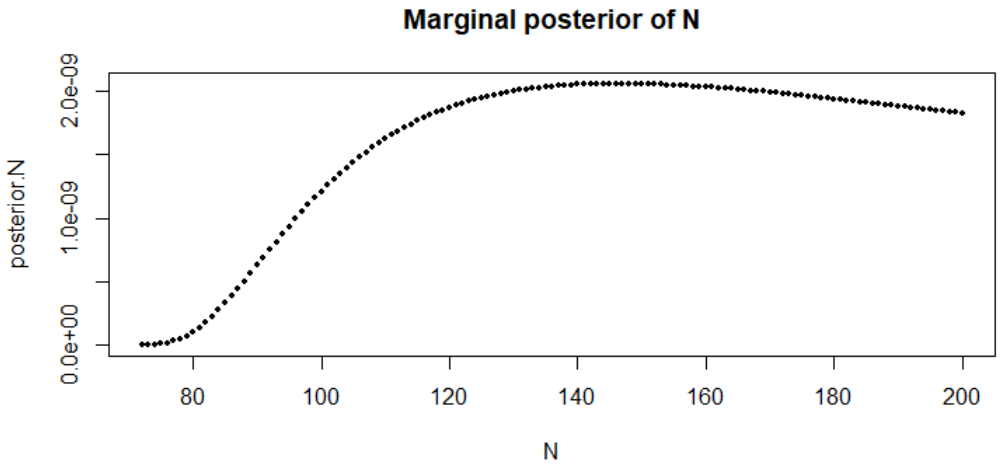
My simulation attempt at this problem isn't perfect, but I tried to simulate this using the following `R` code.

```
waterbuck_data <- c(53, 57, 66, 67, 72)

p.N.data <- function(N, waterbuck_data) {
    return(prod(sapply(waterbuck_data, function(x) choose(N,
        x)))*(beta(sum(waterbuck_data) + 1, 5*N -
        sum(waterbuck_data) + 1)))
}

N <- seq(72, 200, by = 1)
posterior.N <- sapply(N, function(x) p.N.data(x, waterbuck_data))
plot(N,
    posterior.N,
    main = "Marginal posterior of N",
    pch = 19,
    cex = .5)
N[which.max(posterior.N)]
```

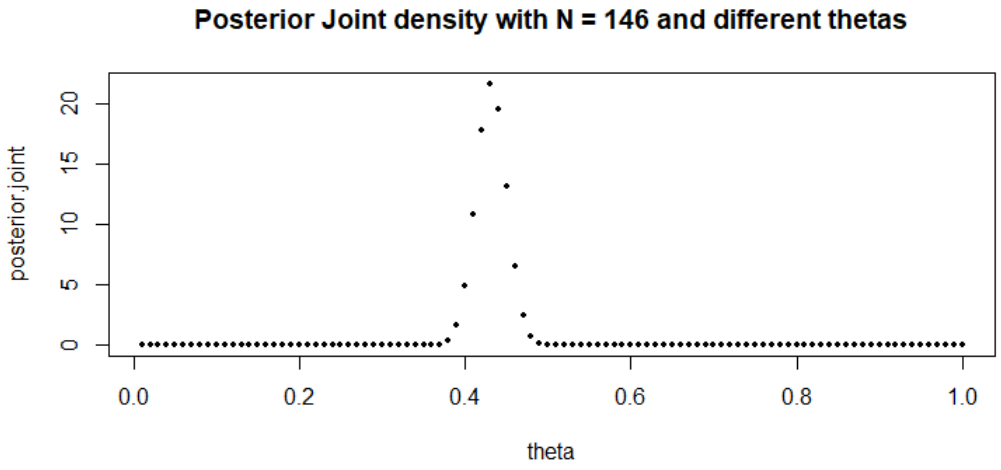
The plot of the marginal posterior of N produced by this code is shown below:



The maximum value of this posterior distribution is $N = 146$ so I used this to create a plot of the posterior joint distribution. Again, this posterior joint is a Beta distribution with $\alpha = \left(\sum_{i=1}^{5} y_i\right) + 1$ and $\beta = \left(5N - \sum_{i=1}^{5} y_i\right) + 1$. So for this problem, I chose to plot this Beta distribution with $N = 146$ and different values of θ.

```
theta <- seq(.01, 1, by = .01)
posterior.joint <- dbeta(theta,
                shape1 = (sum(waterbuck_data) + 1),
                shape2 = (146*5 - sum(waterbuck_data) + 1))
plot(theta,
    posterior.joint,
    main = "Posterior Joint density with N = 146 and different
        thetas",
    pch = 19,
    cex = .5)
theta[which.max(posterior.joint)]
```

This produces the following plot



which has a maximum at $\theta = .43$.

In order to calculate the probability that $N > 100$, I would need a function that gave me the normalized posterior density of N. I was unsure how to code this function in `R` since I did not have an idea of what the normalizing constant would be, but if I did, running the following line should give me the answer that I want:

```
sum(posterior.N[(N > 100)])
```

I approximate this value to be around .933 using the following rough calculation in R where I divide that sum by the total posterior density that I was able to calculate.

```
sum(posterior.N[(N > 100)])/sum(posterior.N)
```

3.3 part c.

Why not simply use a Poisson with fixed μ as a prior distribution for N?

answer The issue with this approach is that we would probably not know which μ to pick beforehand, unless we had some kind of pilot study done beforehand. If we were to pick a μ value from the data, this would be a form of data snooping and invalidate our results.

4 Q10

For $i = 1, 2$, suppose that

$$y_{i1}, \ldots, y_{in_i} | \mu_i, \sigma_i^2 \sim N(\mu_i, \sigma_i^2)$$
$$p(\mu_i, \sigma_i^2) \propto \sigma_i^2$$

and (μ_1, σ_1^2) are independent of (μ_2, σ_2^2) in the prior distribution. Show that the posterior distribution $\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}$ is F with $(n_1 - 1, n_2 - 1)$ degrees of freedom.

answer Let's make use of a previous fact used in the textbook on page 65. For normal i.i.d. random variables, we have the following posterior distribution.

$$p(\sigma^2|data) \propto \left(\sigma^2\right)^{\frac{n-1}{2}-1} exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)$$

Let's also make use of the fact that an F random variable is formed from the ratio of two χ^2 random variables divided by their respective degrees of freedom.

Using a transformation $\lambda = \frac{1}{\sigma^2}$, we see that

$$\begin{aligned} p(\lambda|data) &\propto \left(\frac{1}{\lambda}\right)^{-\frac{n+1}{2}} exp\left(-\frac{\lambda s^2(n-1)}{2}\right)|\frac{1}{\lambda^2}| \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}-\frac{3}{2}} exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \\ &\propto \left(\frac{(n-1)s^2}{\sigma^2}\right)^{\frac{n-1}{2}-1} exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \end{aligned}$$

This is a useful result, since the pdf for a χ^2 distribution with k d.f. is $f(x) \propto x^{\frac{k}{2}-1}exp\left(-\frac{x}{2}\right)$. Therefore, we have the familiar identity that $\frac{(n_1-1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ and $\frac{(n_2-1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$.

Now we can derive the final bit:

$$\begin{aligned} &\left(\frac{(n_1-1)s_1^2}{\sigma_1^2}/(n_1-1)\right) / \left(\frac{(n_2-1)s_2^2}{\sigma_2^2}/(n_2-1)\right) \\ &= \frac{s_1^2}{\sigma_1^2}/\frac{s_2^2}{\sigma_2^2} \\ &= \frac{s_1^2}{s_2^2}/\frac{\sigma_1^2}{\sigma_2^2} \sim F_{n_1-1,n_2-1} \end{aligned}$$